2017

Statistical methods for bullet matching

Eric Riemer Hare
Iowa State University

Statistical methods for bullet matching

by

Eric Riemer Hare

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Co-majors: Statistics; Computer Science

Program of Study Committee:
Heike Hofmann, Co-major Professor
Kris De Brabanter, Co-major Professor
Ulrike Genschel
Jin Tian
Sigurdur Olafsson

The student author and the program of study committee are solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Although I can't possibly acknowledge everyone who has had a positive impact on my academic career, I would like to give a few special shoutouts.

First, to the men and women at the TIBCO Software Seattle office, who offered me the internship which led me to discover my love of statistical computing.

Second, to the Iowa State Department of Statistics, who believed in my ability to conduct research and earn a PhD before I did, and were the first to give me the opportunity.

Third, to my parents Elizabeth and Edward, who despite having PhDs of their own, never pressured or expected me to follow in their footsteps, but supported me all along.

Fourth, to my girlfriend Rainie, who ventured with me to the frozen midwest in pursuit of a PhD of her own, helping to make all this possible.

Fifth, to all my friends I've met in Ames, who helped ease my transition to Iowa State and make me immediately feel comfortable and happy here.

Sixth, to my friends back home in Seattle who have stuck by me as I left them for five years to do research.

Last but absolutely not least, my major professor, Dr. Heike Hofmann, who has been the best mentor I could have hoped to find.

ABSTRACT

Despite being an accepted and established forensic science practice, the process of matching bullets to determine whether they were fired from the same gun barrel has come under fairly intense scrutiny in recent years. This began in earnest in 2009 with a National Academy of Sciences report questioning the scientific validity of these methods. Further criticisms were made in a 2016 report by the President's Council of Advisors on Science and Technology (PCAST). After PCAST determined that there has only been one appropriately designed study to assess the accuracy of bullet matching methods, the report concluded that "[T]he current evidence falls short of the scientific criteria for foundational validity." The report also outlines a way forward by noting "A second—and more important—direction is . . . to convert firearms analysis from a subjective method to an objective method." This thesis attempts to take steps towards that goal. It begins by describing an automatic algorithm for matching bullet lands, and assesses this algorithm on the James Hamby study data. These ideas are then generalized in order to increase the prediction accuracy, determine operator effects in bullet scanning, handle the case of bullet land degradation, and apply to full bullet matches. Finally, a modern web-based database and software system for bullet matching is introduced, allowing for more seamless collaboration in the research community for assessing and improving these algorithms.

CHAPTER 1. INTRODUCTION

1.1 Background

There are few stakes higher than the prosecution of suspects in the criminal justice system. In the United States, 32 states continue to maintain the use of the capital punishment, in which a convicted suspect can have their life ended. All 50 states will consider a lifetime imprisonment, with no opportunity for parole, given the extent of the crime committed. And yet, in spite of the major stakes, many forensic science methods have come under fire in recent years due to a lack of statistical rigor. The issues were well summarized in a report by the President's Council of Advisors on Science on Technology (PCAST) (Advisors on Science and Technology 2016). A nonexhaustive list of the issues:

- Fingerprint analysis is a subjective process where investigators can be swayed, and error rates are seriously lacking.
- Bitemark analysis is seen as a method that probably can't be developed into a scientifically valid method, and resources towards such efforts should be minimal.
- DNA analysis, while more sound, sometimes discounts the role of operator error in the process.

There was also a heavy focus on the limitations of firearms analysis. Though a widely used and accepted procedure, it has come under particular scrutiny in the past decade. In 2005, in *United States vs. Green*, the court ruled that the forensic expert could not confirm that the bullet casings came from a specific weapon with certainty, but could merely "describe" other casings which are similar. Further court cases in the late 2000s expressed caution about the use of firearms identification evidence (Giannelli 2011). In 2009, the National Academy of Sciences published

a report (National Research Council 2009) questioning the scientific validity of many forensic methods including firearm examination. The report states that "[m]uch forensic evidence – including, for example, bite marks and firearm and toolmark identification is introduced in criminal trials without any meaningful scientific validation, determination of error rates, or reliability testing to explain the limits of the discipline." The PCAST report corroborated these findings, and explained how modern techniques could potentially be used to turn this analysis into something more objective:

A second—and more important—direction is (as with latent print analysis) to convert firearms analysis from a subjective method to an objective method. This would involve developing and testing image-analysis algorithms for comparing the similarity of tool marks on bullets. There have already been encouraging steps toward this goal. Recent efforts to characterize 3D images of bullets have used statistical and machine learning methods to construct a quantitative "signature" for each bullet that can be used for comparisons across samples.

To begin to develop such an objective method, the currently available computing tools must be explored and discussed so that they can be leveraged for this purpose.

1.2 Designing a Modern Software System

Conveniently, the computing revolution has opened up statistical methods and tools to a broad range of fields, and these tools can be used to begin such an image analysis algorithm. With the growing popularity of R (R Core Team 2016) in particular, the wide range of choices of open source statistical routines in the form of packages has significantly expanded statistical computing capabilities.

Still, there remains a fundamental obstacle to the use of R. Effective use of R requires a commitment to learning and understanding programming, which some in the forensic science community may not have the desire to do so. Furthermore, although the open-source nature of R is one of its biggest assets, it also means that there is a more rapid development cycle than

would often be found in more corporate software solutions. This means that R developers must continue to maintain their code while learning new programming concepts.

A number of tools have been developed in an attempt to address this issue. The commercial software on which R is derived, S-PLUS (*S-PLUS Programmer's Manual* 1992), includes a rudimentary graphical user interface (GUI) supporting data editing, graphing, and basic statistics. Over time, GUIs were developed for R as well. One of the first was R Commander (Fox 2005), which provides a wrap-around user interface for R. With drop-down menus allowing point-and-click selection of a number of common data analysis and statistical functions, analysis could be performed without a knowledge of programming. More recently, the program Deducer (Fellows 2012) also abstracts the programming into graphical menus and buttons. It expands on R Commander by providing an effective data viewer, help system, and easy to read tables displaying the results.

GUIs have some natural limitations that often make them a less appealing option for researchers. The results of an analysis from a GUI are not typically reproducible. Whereas an R script can be created, shared, and executed elsewhere, the actions taken in a GUI are not transcribed and portable. GUIs also tend to slow down the development and iteration process once the user has become more comfortable with the programming concepts. For instance, scripts allow copying and pasting of code blocks that need only minor modifications. In a GUI, the options representing a code block would need to be individually chosen through drop-down menus.

Recognizing some of these limitations, other approaches have been taken to easing the transition to working with R. RStudio (RStudio Team 2015) provides a GUI around R with expanded functionality, but maintains focus on the scripting and coding aspect. In this sense, RStudio more readily resembles an IDE (Integrated Development Environment), which aid the programmer rather than attempting to abstract the programming away. While this allows reproducibility and may still help a less experienced programmer begin to get started in a programming language, it still depends on a continuing effort to learn programming.

The Shiny package for R provides a framework for researchers with at least a limited knowledge of R programming to make their analysis available to other researchers. Shiny (Chang et al. 2015) is a web development framework which can help turn the results of an R analysis into an interactive web application. Results can be generated by browsing to the website at which the Shiny application is deployed, and using GUI elements (drop-downs, text boxes, tabs) similar to R Commander or Deducer in order to generate results. But a Shiny application is standard R code, and hence maintains the reproducibility and maintainability benefits of standard R scripts.

Because Shiny offers a solution which maintains the benefits of both GUIs and standard programming, I believe it can form the basis for a new set of tools and concepts that greatly expand the reach of statistics. Those who are comfortable with programming can now provide functionality to those who aren't. This functionality can enable researchers to see, understand, and work with their data in ways that they were simply unable to. Ultimately, an open-source solution based on R and Shiny I believe can yield a bullet matching framework which allows for iteration and improvement, but doesn't shut out individuals lacking a knowledge of or a desire to learn programming. Specifically, a modern bullet matching application should aim to be:

- **Modular** - Components of the application can be dynamically enabled and disabled at run-time, allowing flexibility in terms of what functionality is presented to the end-user.
- **Extensible** - New components, or modules, can be written to further extend the functionality beyond what the base application allows.
- **Web-Based** - The application should live on the web, so that it can be accessed anywhere, from any device, without the need to deal with restrictive licenses or unsupported platforms.
- **Reproducible** - Results generated by the application should be reproducible. There should be no black boxes.

These ideas are described further over the course of this document in order to develop a modern, reproducible, and statistical application for the analysis of bullet lands. Before arriving at that point, however, we must further motivate the need for modern bullet matching algorithms.

1.3 Algorithms for Bullet Matching

In the United States, suspects are considered innocent until proven guilty "beyond a reasonable doubt". This in many ways parallels traditional hypothesis testing approaches, in which a pre-defined cut-off (significance level) is used to determine the threshold at which the null hypothesis is rejected (which presumably should occur once the evidence leads us beyond a reasonable doubt).

Rifling, manufacturing defects, and impurities in a barrel create striation marks on the bullet during the firing process. These marks are assumed to be unique to the barrel, as described in a 1992 AFTE article (AFTE Criteria for Identification Committee 1992). Current standard practice for bullet matching relies in part on the assessment of the so-called maximum number of consecutively matching striae (CMS), first defined by Biasotti (1959). One of the primary issues with this procedure is that a human inspection to determine CMS is subjective (Miller 1998). Human inspection also requires on-site analysis of the bullets, which can be costly and time-consuming, and introduces the potential for differing opinions across different forensic examiners.

A modern development in this realm is the adoption of an open format for storing 3D topographical images of bullets in a format called x3p (XML 3-D Surface Profile). The x3p format conforms to the ISO5436-2 standard[1], implemented to provide a simple and standard conforming way to exchange 2D and 3D profile data. It was adopted by the OpenFMC (Open Forensic Metrology Consortium[2]), a group of academic, industry, and government firearm forensics researchers whose aim is to establish best practices for researchers using metrology in forensic science. Furthermore, NIST (the National Institute for Standards and Technology) is developing a database to allow searching and downloading of these x3p files[3]. Although limited to around 70 bullets at the time of this writing, this database in conjuction with open-source software to

[1] http://sourceforge.net/p/open-gps/mwiki/X3p/
[2] http://www.openfmc.org/
[3] https://tsapps.nist.gov/NRBTD/

work with .x3p files opens up a whole new set of possibilities in terms of a statistical foundation for bullet matching.

The feasibility of creating a database of ballistic images that could be used to identify guns used in crimes was evaluated in a 2008 report by the National Research Council (Committee to Assess the Feasibility, Accuracy and Technical Capability of a National Ballistics Database 2008). The evaluation investigated the scalability of NIBIN (National Integrated Ballistic Information Network), which uses proprietary matching algorithms provided by IBIS. The bottom line of the report was that in spite of the many technical and practical hurdles, solutions to all but one problem could be found. The problem that remained is that statistically, the quality of the matching algorithm (in this case, of breech-face marks and firing pin impressions) could not withstand a hugely increased number of records while still maintaining a reasonable workload for forensic examiners, who have to examine possible matches suggested by the system.

We have several broad goals in developing a modern statistical matching algorithm. First, we wish to define every statistic or measure used objectively. Second, we will make the definitions and code open-source and publicly accessible so that its open for review by forensic scientists and statisticians alike. Third, we will investigate the distributional properties of these statistics across the available universe of bullets accessible to us in the database. Finally, we wish to provide an easy-to-use interface to serve as a front-end for the algorithm.

Critical to the success of a matching algorithm is the extraction of a set of features describing a bullet signature. In addition to the aforementioned CMS, the CCF is used, as it has for other bullet matching applications (Vorburger et al. 2011). Traditional bullet matching methods have used strict cutoffs (for instance, 6 CMS) to determine a match versus a non-match. We are aiming to be more robust in using a number of features and deriving conditional probabilities of matches given particular values of these features.

- *CCF*: Function of the optimum shift distance measuring the correlation between two profiles (Vorburger et al. 2011)

- *CMS*: Striated markings that line up exactly with one another without a break or dissimilarity in between them (Biasotti 1959, Chu et al. (2013)). This and other forensic science papers using CMS typically count a single peak as a striae, while we count peaks and valleys, so our definition typically yields CMS values about twice those commonly found in the literature.
- *CMNS*: Striated markings that do not line up exactly with another, without matching striation between them.
- *Matches*: The number of matching striations between two signatures
- *Non-Matches*: The number of unmatched striations between two signatures
- $D = \sqrt{\frac{1}{\#t}\sum_t [f(t) - g(t)]^2}$ where $f(t)$ and $g(t)$ are aligned signatures. The euclidean vertical distance between surface measurements of aligned signatures. This is a measure of the total variation between two profiles (Clarkson and Adams 1933).
- S: The sum S of average absolute heights of matched extrema: for each of the two matched stria, compute the average of the absolute heights of the peaks or valleys. S is then defined as the sum of all these averages.

With this in mind, we have developed an automated matching routine, written in R, which uses open and transparent statistical techniques to arrive at a predicted probability of a match at the bullet land level. This framework is provided as an R package called `bulletr` (Hofmann and Hare 2016) with an associated web component, discussed in the next section. This is not the first automatic bullet matching system (Xie et al. 2009, Riva and Champod (2014), Bachrach (2002)). But it builds on strong research principles by using a publicly accessible database, including fully reproducible results, and using a broad set of derived features to produce probabilities or scores based on a machine learning algorithm.

This work has been submitted and accepted (with revisions) by the Annals of Applied Statistics. We are following up by investigating the properties of different features as applied to degraded bullets, and when compared with common cut-offs for match, non-match, and inconclusive from the literature. In particular, because a real world scenario often involves recovering only a fragment of the bullet from a crime scene, many of the traditional features such as CMS

need generalizations that handle these cases. Furthermore, there is an open question regarding precisely the size of the fragment needed in order to be confident of a match "beyond a reasonable doubt". This warrants further investigation.

1.4 A Web Framework for Bullet Matching

Critical to the success of a software system like this is that it can be used, extended, and enhanced by a broad range of scientists. This means we need a reproducible web-based software system that opens the R programming tools used for bullet matching to forensic examiners and forensic scientists who may not have the knowledge of, or desire to learn, programming. There are three primary components to this software system: the database, the front-end application, and the back-end application.

The first component is a database. Our database builds on top of the NIST Ballistics Research Database[4]. NIST's database provides an open and transparent source of raw data files representing surface topologies of toolmarks. Our database allows the storage of **processed** versions of these toolmark surfaces, tracking the parameters necessary to reproduce the results. This allows for a quantitative assessment of each step of our bullet matching algorithm. The database is discussed at length in Chapter Four.

The second component of our software system is a web front-end aimed at those who do not have R programming skills. The web front-end allows forensic examiners to upload bullet land images, examine the surface topologies, and perform each aspect of the algorithm in order to arrive at a probability of a match. The final display includes a results page in which all chosen parameters of the algorithm are provided so that a report on the results can be generated and cross-checked by other researchers. Figure 1.1 displays a screenshot of this component.

[4] https://tsapps.nist.gov/NRBTD

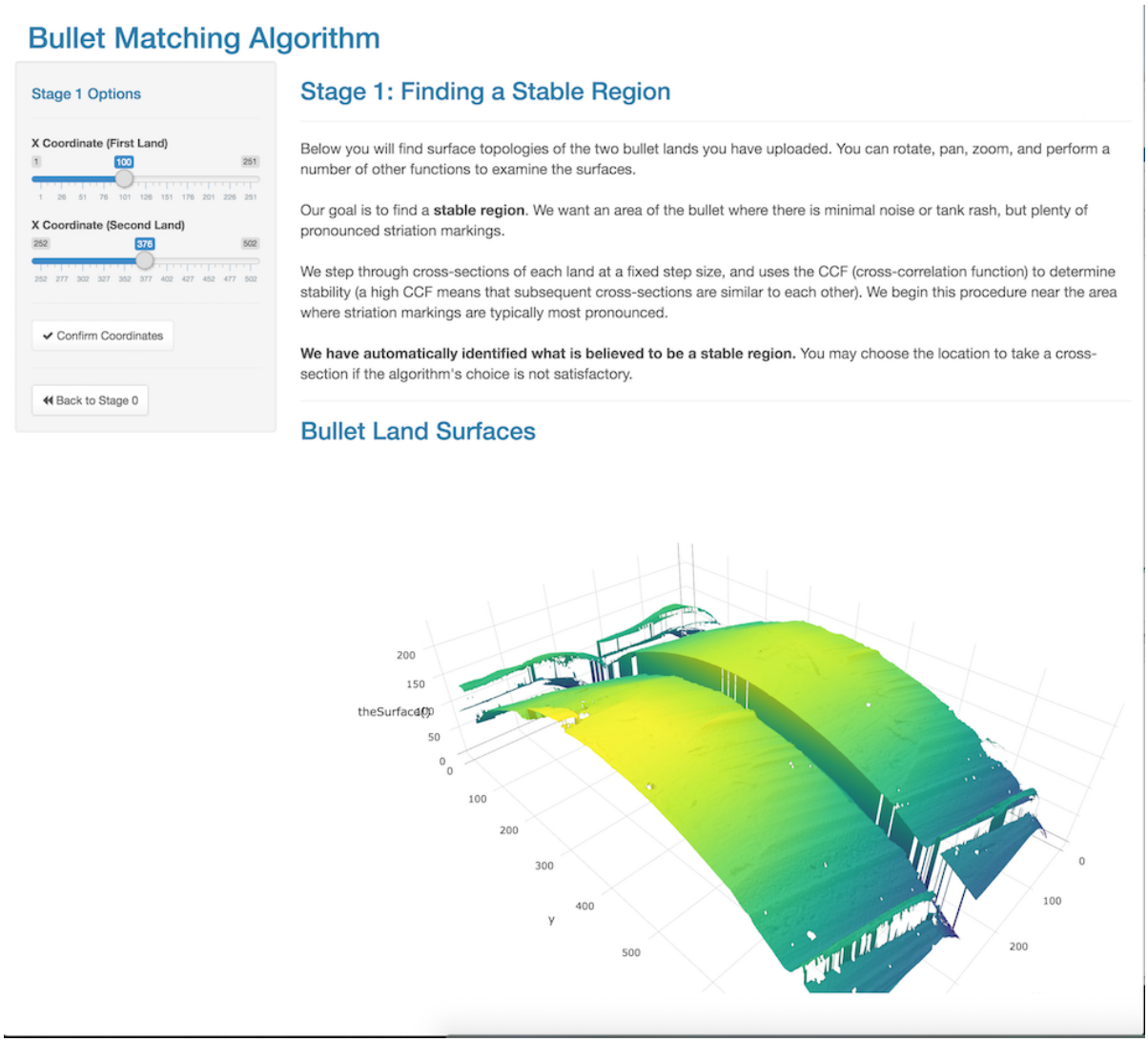


Figure 1.1: Prototype user interface for the bullet matching algorithm.

Finally, the third component is a web-based backend designed for the further assessment and development of the algorithms itself. This is aimed at researchers and scientists who wish to iterate on the performance of the algorithm by continuing to assess its results in the context of the database, and the different processed versions of toolmark images available.

Together, these components form the basis of a modern software system that implements the algorithms from Chapters Two and Three using the software design principles I've outlined. Specifically, this system is **modular** in that the different components can be substituted as

needed to fit new use cases. The system is also **extensible**, as functionality can be added to the R package which immediately gets reflected by the web-based systems. It also is fully **web-based**, requiring just a web browser to access the database. And finally, the results are all **reproducible** as each parameter yielding intermediate results is tracked throughout, and available for viewing in the database at any time.

CHAPTER 2. AUTOMATIC MATCHING OF BULLET LANDS

A paper submitted to the **Annals of Applied Statistics**.

Eric Hare, Heike Hofmann, Alicia Carriquiry

Abstract

In 2009, the National Academy of Sciences published a report questioning the scientific validity of many forensic methods including firearm examination. Firearm examination is a forensic tool used to help the court determine whether two bullets were fired from the same gun barrel. During the firing process, rifling, manufacturing defects, and impurities in the barrel create striation marks on the bullet. Identifying these striation markings in an attempt to match two bullets is one of the primary goals of firearm examination. We propose an automated framework for the analysis of the 3D surface measurements of bullet lands that first transcribes the markings into a 2D plotting framework. This makes identification of matches easier and allows for a quantification of both matches and matchability of barrels. The automatic matching routine we propose manages to (a) correctly identify lands (the surface between two bullet grooves) with too much damage to be suitable for comparison, and (b) correctly identify all 10,384 land-to-land matches of the James Hamby study (Hamby, Brundage, and Thorpe 2009).

2.1 Introduction

Firearm examination is a forensic tool used to help the court determine whether two bullets were fired from the same gun barrel. This process has broad applicability in terms of convictions in the United States criminal justice system. Firearms identification has long been considered an accepted and reliable procedure, but in the past ten years has undergone more significant scrutiny. In 2005, in *United States vs. Green*, the court ruled that the forensic expert could not confirm that the bullet casings came from a specific weapon with certainty, but could merely "describe" other casings which are similar. Further court cases in the late 2000s expressed caution about the use of firearms identification evidence (Giannelli 2011).

In 2009, the National Academy of Sciences published a report (National Research Council 2009) questioning the scientific validity of many forensic methods including firearm examination. The report states that "[m]uch forensic evidence – including, for example, bite marks and firearm and toolmark identification is introduced in criminal trials without any meaningful scientific validation, determination of error rates, or reliability testing to explain the limits of the discipline."

Rifling, manufacturing defects, and impurities in a barrel create striation marks on the bullet during the firing process. These marks are assumed to be unique to the barrel, as described in a 1992 AFTE article (AFTE Criteria for Identification Committee 1992). "The theory of identification as it pertains to the comparison of toolmarks enables opinions of common origin to be made when the *unique surface contours* of two toolmarks are in sufficient agreement". The article goes on to state that "Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges and furrows."

From a statistical standpoint, identification of the gun that fired the bullet(s) requires that we compare the probabilities of observing matching striae under the competing hypotheses that the gun fired, or did not fire, the crime scene bullet. If indeed the uniqueness assumption is plausible, the latter probability approaches zero.

Current firearm examination practice relies mostly on visual assessment and comparison of striation. Indeed, the AFTE Theory of Identification (https://afte.org/about-us/what-is-afte/afte-theory-of-identification) explicitly requires that examiners evaluate the strength of similarity between two samples relative to other comparisons they may have carried out in the past. An attempt to quantify the degree of similarity consists in counting the number of consecutively matching striae (CMS) between two bullets, first proposed by Biasotti (1959). This approach has two drawbacks, however. First, determining matching striae is still a subjective activity. Second, as discussed by Miller (1998), the number of CMS may be high even if the bullets were not fired by the same gun.

Here, we focus on the question of defining a metric that can be used to objectively compare two bullets. We propose a framework which allows for the automatic analysis of the surface topologies of bullets, and the transcription of the individual characteristics into a 2D plotting framework.

We work with images from the James Hamby Consecutively Rifled Ruger Barrel Study (Hamby, Brundage, and Thorpe 2009). Ten consecutively rifled Ruger P-85 pistol barrels were obtained from the manufacturer and fired to produce 20 known test bullets and 15 unknown bullets for comparison. 3D topographical images of each bullet were obtained using a NanoFocus lens at 20x magnification and made publicly available on the NIST Ballistics Database Project[5] in a format called x3p (XML 3-D Surface Profile). The x3p format conforms to the ISO5436-2 standard[6], implemented to provide a simple and standard conforming way to exchange 2D and 3D profile data. It was adopted by the OpenFMC (Open Forensic Metrology Consortium[7]), a group of academic, industry, and government firearm forensics researchers whose aim is to establish best practices for researchers using metrology in forensic science. We have developed an open-source package for analyzing bullet lands written in R (R Core Team 2016). This package is called bulletr (Hofmann and Hare 2016) and enables direct reading and manipulation of x3p files. It also implements all of the methods we propose in this paper. A different package

[5] http://www.nist.gov/forensics/ballisticsdb/hamby-consecutively-rifled-barrels.cfm
[6] http://sourceforge.net/p/open-gps/mwiki/X3p/
[7] http://www.openfmc.org/

exists for reading x3p files called x3pr (OpenFMC 2014) developed by Petraco (2014), but it is not designed to carry out calculations like the ones we propose after the x3p files have been read.

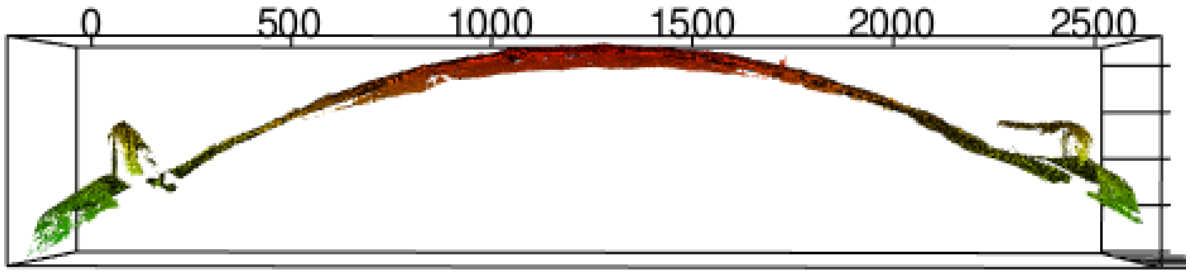


Figure 2.2: View of the data along the circumference of the bullet (circular segment of about 30 degrees).

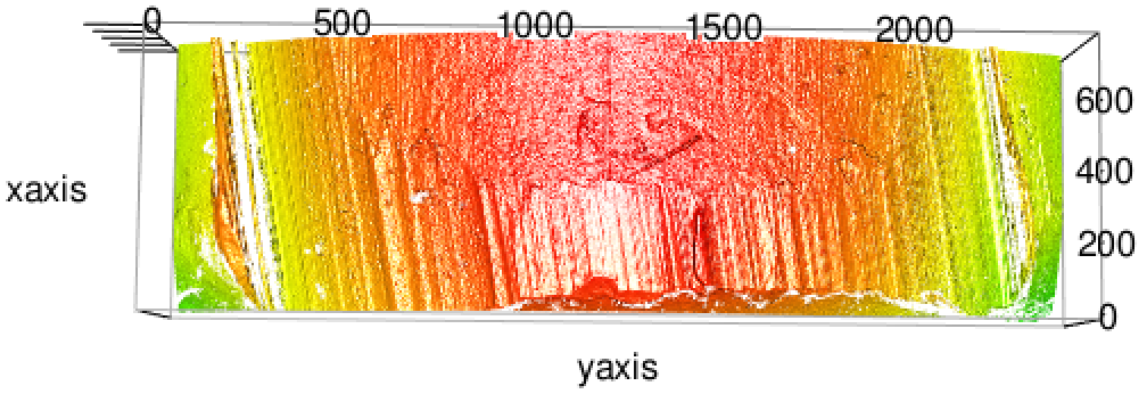


Figure 2.3: Frontal view of a bullet land (lower end of the view is the bottom of the bullet).

Each fired bullet is provided in the form of a set of six x3p files, where each file is a surface scan between adjacent grooves on the bullet, called a "land". In the Hamby data, typical length (groove-to-groove) of a land is about 2005.9 micrometers or 2.01 millimeters. For notational simplicity, we refer to a particular land of a bullet as bullet X-Y, where X is the bullet identifier, and Y is the land number. An example of plotting one of these lands is given in Figures 2.2 and 2.3. These figures show side and top profiles of the land respectively. The tilt of the lines to the left in Figure 2.3 is not an artifact, but a direct and expected consequence of the spin induced by the rifling during the firing process. Depending on whether a barrel is rifled clockwise or

counter-clockwise, the striations have a left or right tilt. The direction of the rifling is a class characteristic, i.e. a feature that pertains to a particular class of firearms, and is not unique at the individual barrel/bullet level.

The typical number and width of striation markings on bullets varies significantly depending on the gun barrel. For instance, a Smith and Wesson barrel with a land-width of 2.4 millimeters contained an average 60 striae, with an average width of about 0.08 millimeters (Chu et al. 2011).

The purpose of our paper is to present an automatic matching routine that allows for a completely objective assessment of the strength of a match between two bullet lands. While we assess the performance of the algorithm in terms of a binary decision of match vs. non-match using a 50% probability cut-off, our primary goal is to highlight the features that are statistically associated with matches and non-matches, and to provide a quantitative assessment of this association. In a real-world application of our algorithm, the raw scores would need further analysis and scrutiny, and it is likely a 50% cut-off would be an inappropriate choice on the basis of reasonable doubt.

Our algorithm is fully open source and available on GitHub (Hofmann and Hare 2016). This transparency allows for a greater understanding of the individual steps involved in the bullet matching process, and allows other forensic examiners, as well as outside observers, to examine the factors that discriminate between known bullet matches and non-matches. We have chosen to perform the matching on a land-to-land level, rather than bullet-to-bullet level. Although doing so introduces an implicit assumption of independence between lands, assuming independence only serves to make the task more challenging.

The remainder of this paper is structured as follows: We first briefly review some earlier work. We then discuss two methods of modeling the class structure of the bullet surfaces. Finally, we proceed to describing an automatic matching routine which we evaluate on the bullets made available through the Hamby study.

2.2 Previous Work

There have been attempts to develop automatic or semi-automatic matching protocols, but most have focused on breech face and firing pin marks (e.g., Riva and Champod 2014) or discuss a single attribute for comparison (e.g., Vorburger et al. 2011, Chu et al. (2011)). Still others refer to proprietary algorithms (Roberge and Beauchamp 2006). We briefly review some of this earlier work in what follows.

The original paper on the complete Hamby study already reports the successful use of several computer-assisted methods. However, aside from a zero false positive rate, false-negative error rates for bullets are not given nor are error rates for land-to-land matches mentioned.

Lock and Morris (2013) proposed an approach to quantify similarity of toolmarks. Their algorithm determines an optimal matching window between two toolmark signatures, and then performs a set of both coordinated and independent shifts. Given a match, the coordinated shifts would be expected to yield correlation values higher than those obtained from independent shifts. This is assessed using a Mann-Whitney U Statistic.

A procedure for bullet matching using the BulletTrax3D system is described in Roberge and Beauchamp (2006). Their study used a different set of ten consecutively rifled barrels; matches are identified based on a bullet-to-bullet correlation score. The authors state that this process 'could be automated', but no implementation of the algorithm is available.

Modern automated techniques using 3D images have also been proposed by e.g. Riva and Champod (2014). However, the authors focused on cartridge cases and not bullets. This might seem like a trivial distinction, but it has implications for the development of the algorithm. The algorithm performs alignment of striae by rotation of the XY plane, which is not generalizable to bullets in which the XY plane is not flat.

Other work on 3D images has been described by Petraco and Chan (2012), who also focus on cartridge cases, as well as screwdriver striation patterns, and by others (e.g., Chu et al. 2011, Chu et al. (2010), Vorburger et al. (2011)).

2.3 Bullet Signatures

To analyze the striation pattern, we extract a **bullet profile** (Ma et al. 2004) by taking a cross section of the surface measurements at a fixed height x along the bullet land. Figure 2.4 shows a plot of the side profile of a bullet land. It can be seen that the global structure of the land dominates the appearance of the plot. The grooves can be clearly identified on the left and right side, and the curvature of the surface is the most visible feature in the middle.

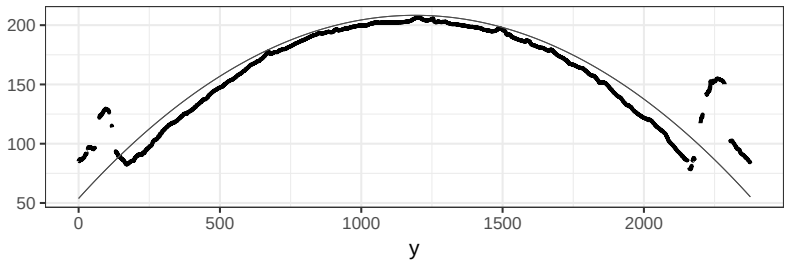


Figure 2.4: Side profile of the surface measurements (in μm) of a bullet land at a fixed height x. Note that the global features dominate any deviations, corresponding to the individual characteristics of striation marks.

The smooth curve on the plot represents a segment of a perfect circle with the same radius as the bullet. While the circle is an obvious first choice for fitting the structure, it does not completely capture the bullet surface after it was fired. A discussion of a circular fit and the remaining residual structure can be found in Section 2.8.1.

Instead of a circular fit, we use multiple loess fits to model the overall structure and extract the bullet markings.

2.3.1 Identifying Groove Locations

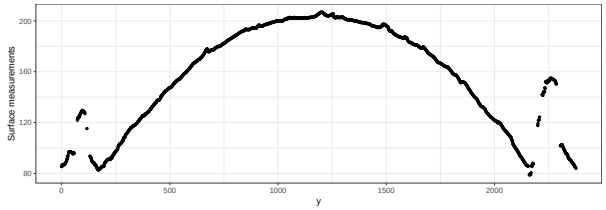
We first identify the location of the left and right grooves in the image. The grooves are assumed to contain no information relevant for determining matches. They also dominate the structure, and therefore need to be removed.

Fortunately, the location and appearance of the grooves in the surface profiles is quite consistent. Surface measurements reach local maxima around the peak of the groove at either end of the range of y, and we can then follow the descent of the surface measurements inwards to the valley of the groove. The location of the valleys mark the points at which we trim the image. The procedure can be described as follows:

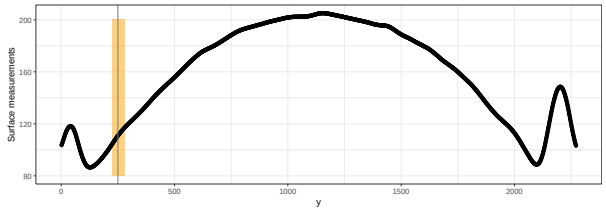
1. At a fixed height x extract a bullet's profile (Figure 2.5a, with $x = 243.75\mu m$).
2. For each y value, smooth out any deviations occurring near the minima by twice applying a rolling average with a pre-set *smoothing factor* s. (Figure 2.5b, smoothing factor $s = 35$ corresponding to $55\mu m$).
3. Determine the location of the peak of the left groove by finding the first doubly-smoothed value y_i that is the maximum within its smoothing window (e.g. such that $y_i > y_{i-1}$ and $y_i > y_{i+1}$, where i is between 1 and $\lfloor s/2 \rfloor$). We call the location of this peak p_ℓ (see Figure 2.5c).
4. Similarly, determine the location of the valley of the left groove by finding the first double-smoothed y_j that is the minimum within its smoothing window. Call the location of this valley v_ℓ.
5. Reverse the order of the y values and repeat the previous two steps to find the peak and valley of the right groove, (p_r, v_r).
6. Trim the surface measurements to values within the two grooves (i.e. remove all records with $y_i < v_\ell$ and $y_i > v_r$) (see Figure 2.5c).

The smoothing factor s introduced in the algorithm represents the window size to use for a rolling average. Higher values of s therefore lead to more smoothing. Empirically, a value of $s = 35$ for the smoothing factor seems to work well (the smoothing factor is further discussed

(a) Step 1 of identifying groove locations: For a fixed height ($x = 243.75\mu m$) surface measurements for bullet 1-5 are plotted across the range of y.



(b) Step 2 of identifying groove locations: The surface measurements are smoothed twice with a smoothing factor of $s = 35$. The orange rectangle shows an example of the smoothing window. Valleys and peaks are detected, if they are not within the same window.



(c) Steps 3 – 6 of identifying groove locations: After smoothing the surface measurements extrema on the left and right are detected (marked by vertical lines, red indicating peaks and blue indicating valleys). Values outside the blue boundaries are removed (shown in grey)

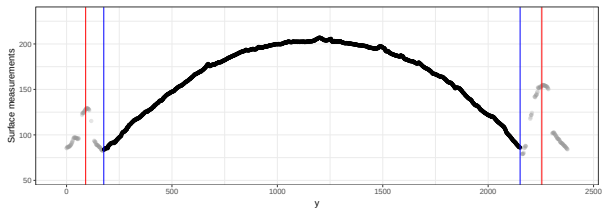


Figure 2.5: Overview of all six steps of the smoothing algorithm to identify and remove grooves from the bullet images.

in Section 2.4.4). It is important to note that the smoothing pass is done twice. That is, the smoothed data are once again smoothed by computing a new rolling average with the same smoothing factor. This bears some similarities to the ideas of John Tukey in his book Exploratory Data Analysis, where he describes a smoothing process called “twicing" in which a second pass is made on the residuals computed from the first pass and then added back to the result (Tukey 1977). This has the effect of introducing a bit more variance back into the smoothed data. We instead performed a second smoothing pass on the smoothed data, which has the effect of weighting observations near the center of the window the highest, with the weights linearly dropping off as we reach either end of the smoothing window.

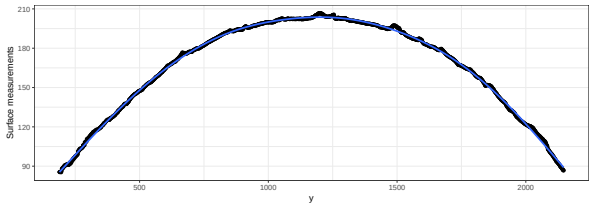
2.3.2 Removing Curvature

Next, we fit a loess regression to the data. Loess regression (Cleveland 1979) is based on the assumption that the relationship between two random variables X and Y can be described in the form of a smooth, continuous function f with $y_i = f(x_i) + \varepsilon_i$ for all values $i = 1, ..., n$. The function f is approximated via locally weighted polynomial regressions. Parameters of the estimation are α, the proportion of all points included in the fit (here, $\alpha = 0.75$), the weighting function and the degree of the polynomial (here, we fit a quadratic regression).

The main idea of locally weighted regression is to use a weighting routine that emphasizes the effect of points close to the fitting location and de-emphasizes the effect of points as they are further away. The weighting function used here is the tricubic function $w(d) = \left(1 - d^3\right)^3$, for $d \in [0, 1]$ and $w(d) = 0$ otherwise. Here, d is defined as the distance between x_i and the location of the fit x_o and the maximum distance of the range of the x-values for span α in x_o.

Figure 2.6a shows the loess fit, in blue, overlaid on the processed image of bullet 1-5. The fit seems to do a reasonable job of capturing the structure of the image. Figure 2.6b shows the residuals from this fit. These residuals are called the **signature** of bullet 1-5.

(a) Loess fit for bullet 1-5.



(b) Residuals of loess fit for bullet 1-5.

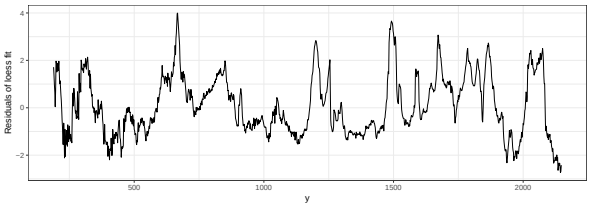


Figure 2.6: Fit and residuals of a loess fit to bullet 1-5 (Barrel 1). The residuals define the *signature* of bullet 1-5.

2.4 Automatic Matching

Applying the loess fit to a range of different signatures (see Figure 2.7 for signatures extracted at heights between $50\mu m$ and $150\mu m$) shows the 3D striation marks from two bullets. Signatures of bullet~1 are shown on the left (all extracted from heights below $100\mu m$) and signatures of bullet 2 are shown on the right (extracted at heights above $100\mu m$). Signatures are manually aligned, resulting in many of the striation marks to continuously pass from one side to the other. Visually, this allows for an easy assessment of these two bullet lands as a match. However, this match relies on visual inspection and is therefore subjective. The goal of this section is to eliminate the need for a visual inspection during the matching process and replace it by an automatic algorithm. This also allows for a quantification of the strength of the match.

In this section, we describe the algorithm for matching signatures first, and the impact of parameter choices in the subsections thereafter.

2.4.1 Algorithm

Figure 2.8 gives an overview of the automated matching routine: We first identify a stable region for each bullet land and extract the signature at the lowest height in this region, because typically, individual characteristics are best expressed at the lower end of the bullet, near the base (see Section 2.8.3 for a more detailed discussion). All of the other steps are done on pairs of bullet lands:

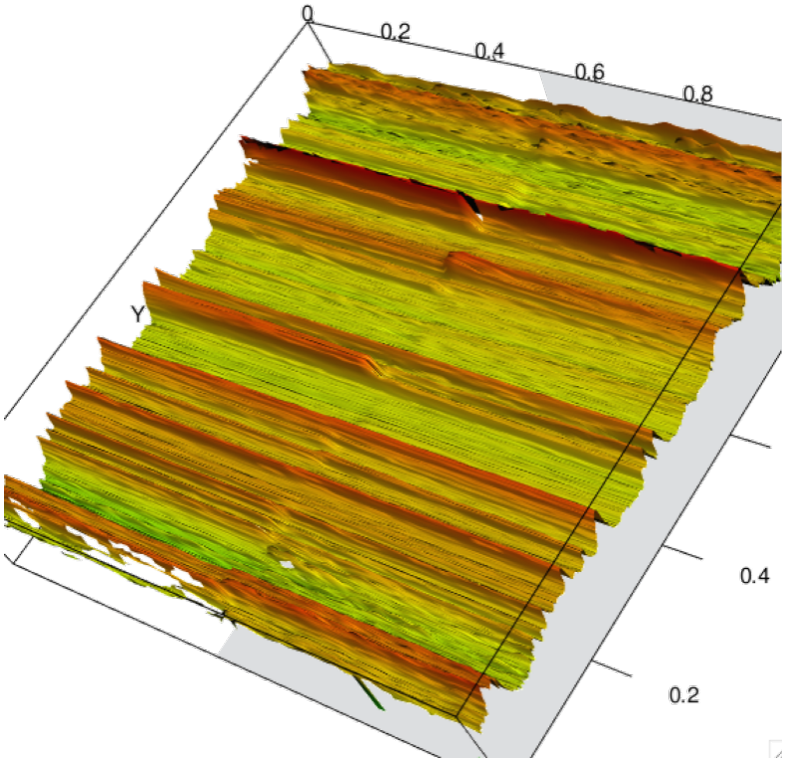
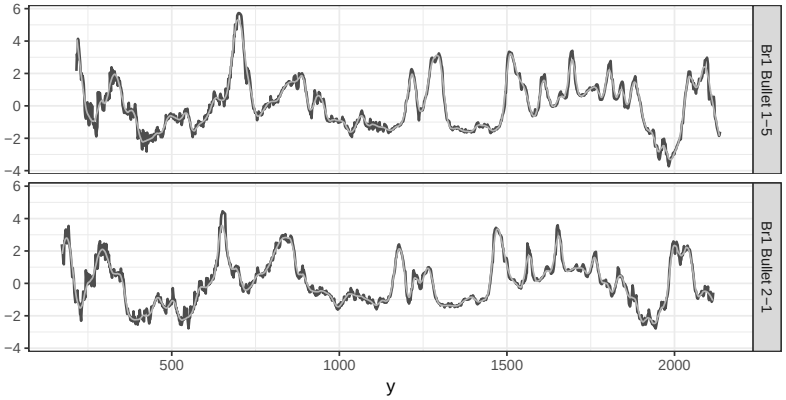
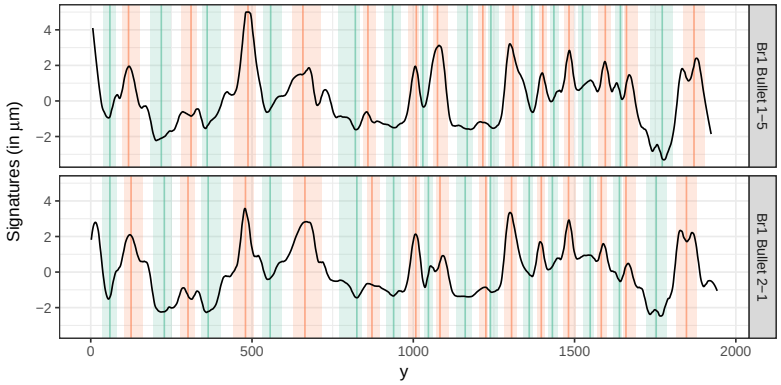


Figure 2.7: 3D view of the manually adjusted side-by-side comparison of bullet 1-5 and bullet 2-1 after removing the curvature. Bullet 2-1 is shaded light grey in the background.

(a) Loess smooth of signatures at a height of $x = 100\mu m$ (span is 0.03).



(b) Using a rolling median peaks and valleys are identified for each signature. Peaks and valleys on the signature correspond to striation marks on the bullet's surface.



(c) Rectangles in the back identify a striation mark on one of the bullets. Matching striation marks are indicated by color filled rectangles and marked by an 'o'. Mismatches are filled in grey and marked by an 'x'.

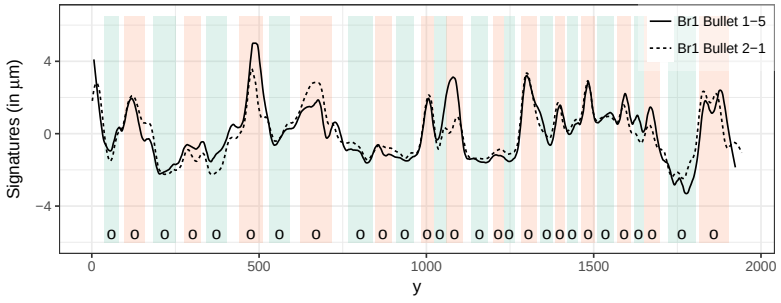


Figure 2.8: Matching striation marks: smooth (a), identify peaks and valley (b), and match peaks and valleys between signatures (c).

1. **Smooth the two signatures** using a loess with a very small span (see Figure 2.8a).
2. Use cross-correlation to **find the best alignment** of the two signatures: shift one of the signatures by the lag indicated by the cross-correlation function (see Figure 2.10 for the cross-correlation function and Figure 2.9b for the resulting shift).
3. Using a rolling average, **identify peaks and valleys** for each of the signatures. We then define an interval around the location of the extrema on each side as one third of the distance to the location of the next extrema (see Figure 2.8b). Peaks and valleys constitute the *striation marks* on the bullet.
4. **Match striations across signatures:** based on the intervals around the extrema as defined above, we identify common intervals as the areas in which two or more of the individual intervals overlap: a joint interval is defined as the smallest interval that encompasses all of the overlapping intervals. A joint interval is then called a match(ing stria) between the signatures, if all of the intervals are of the same type of extrema, i.e. they are either all peaks or all valleys. In Figure 2.8 all matches are shown as color-filled rectangles corresponding to their type of extrema (peaks are shown in orange, and valleys in green). Non-matching intervals are left grey.
5. **Extract features from the aligned signatures and the matches between them:** many different features can be extracted from the aligned signatures. Here, we describe a few of the ones that can be found in the literature and some that we found to be of practical relevance:

   i. Maximal number of CMS (consecutive matching striae), and, similarly, the number of consecutively non-matching striae (CNMS),
   ii. Number of matches and non-matches,
   iii. The value of the cross-correlation function (ccf) between the aligned signatures (Vorburger et al. 2011),
   iv. Average difference D between signatures, defined as the Euclidean vertical distance between surface measurements of aligned signatures. Let $f(t)$ and $g(t)$ be smoothed,

aligned signatures:

$$D^2 = \frac{1}{\#t} \sum_t [f(t) - g(t)]^2 \,,$$

v. The sum S of average absolute heights of matched extrema: for each of the two matched stria, compute the average of the absolute heights of the peaks or valleys. S is then defined as the sum of all these averages.

The difference D between signatures is here defined as the Euclidean distance (in μm). In the paper by Ma et al. (2004), distance is defined as a measure relative to the first signature, which serves as a comparison reference and is therefore a unitless quantity.

Counting the maximal number of CMS is part of the current practice to identify bullet matches (Nichols 1997, Nichols (2003a), Nichols (2003b)). In the example of Figure 2.8, the number of consecutive matching striations (CMS) is fifteen, a high number suggestive of a match between the lands. Note that the definition of CMS we use does not match the one given in Chu et al. (2013). There, CMS is defined only in terms of matching peaks without regarding valleys. Additionally, peaks in Chu et al. (2013) are used only if they can be identified and matched 'within a tolerable range' between lands. The definition given here is computationally less complex, but should yield highly correlated values, because of the requirement to only consider signatures from a stable region in the land (see Section 2.4.3 for further details on stability of regions). In the Hamby study, the definition of CMS by Chu et al. (2013) leads to approximately half of the values of CMS defined in this paper (with a correlation coefficient between the values of the two definitions of about 0.92). For lead bullets, such as used in the Hamby study, Biasotti (1959) considered four or more consecutive peaks (corresponding to eight or more consecutive lines in our definition) to be sufficient evidence of a match.

Determining a threshold such that CMS values above the threshold indicate a match with high reliability is beyond the scope of this work, even though it is critically important in practice. We provide some ideas in the next section, but first we assess the robustness of the matching algorithm to different choices of the parameter values.

2.4.2 Horizontal Alignment

Signatures of each of the two lands, 1-5 and 2-1, in Figure 2.7 are shown in Figure 2.9 extracted at a height of $x = 100\mu m$. Striation marks show up in these representations as peaks and valleys. The individual characteristics are prominent and, again, suggest a match between the lands. A horizontal shift of one of the signatures (result shown in Figure 2.9b) emphasizes the strong similarities between signatures.

(a) Raw bullet land signatures. (b) Aligned signatures.

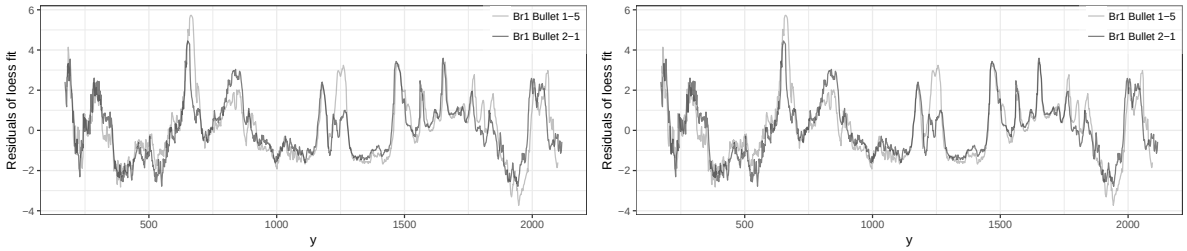


Figure 2.9: Signatures of bullets 1-5 and 2-1 taken at heights of $x = 100\mu m$. A horizontal shift of the values of bullet 1-5 to the right shows the similarity of the striation marks.

For this alignment we use the cross-correlation function to find a maximal amount of agreement between the signatures (Bachrach 2002, Chu et al. (2010), Vorburger et al. (2011), Chu et al. (2013)). This horizontal shift is based on the cross-correlation between the two signatures: let $f(t)$ and $g(t)$ define the signature values at t, where t are locations between 0~μm and about 2500 μm, 1.5625 μm apart. The cross-correlation between f and g at lag k is then defined as

$$(f * g)(k) = \sum_t f(t+k)g(t),$$

with suitably defined limits for the summation.

2.4.3 Impact of Bullet Height

The height at which signatures are extracted for a comparison between bullet lands matters – signatures taken from heights that are further apart, show more pronounced differences between the signatures. This poses both a caveat to matching attempts as well as an opportunity for

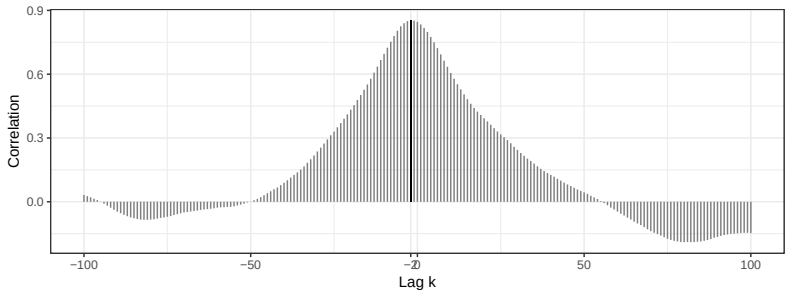


Figure 2.10: Cross-correlation function between the two signatures shown in Figure 2.9a at lags between -100 and 100. The correlation is maximized at a lag of -17, indicating the largest amount of agreement between the signatures. Figure 2.9b shows the lag-shifted signatures.

quality control: we have to be aware of the height that was used in a matching. Visually, matches degrade if the signatures upon which the match is based are from heights further than $200\mu m$ apart (see Section 2.8.2 for more discussion). However, we can extract signatures from multiple heights of the same bullet land for an initial assessment of its quality. By comparing signatures from heights that are not too far apart – $25\mu m$ to $50\mu m$ – we get an indication whether the signatures come from a rapidly changing section of the surface, indicative of a break-off or some other damage, or from a stable section, where we have a reasonable expectation of finding matches to other signatures. In the approach here, we keep increasing the height x at which the signature is taken until we find a section with a stable pattern. This process is shown in Figure 2.11 at the example of bullet 1-1 from barrel 3, where 'stability' is defined as two aligned signatures from heights chosen $25\mu m$ apart having a cross-correlation of at least 0.95.

2.4.4 Varying Smoothing Factor

As mentioned earlier, the algorithm for detecting peaks and valleys depends on the selection of a smoothing window, called the smoothing factor or span. A smoothing factor of k means that the k closest observations to x_o are considered for a fit for x_o. Because surface measurements are recorded at an equidistant resolution (here, of $1.5625\mu m$), we decided to only consider odd smoothing factors $2k+1$, which means that the k observations to the left and right of x_o are considered for a local fit of x_o. For detecting and removing the grooves prior to fitting a loess

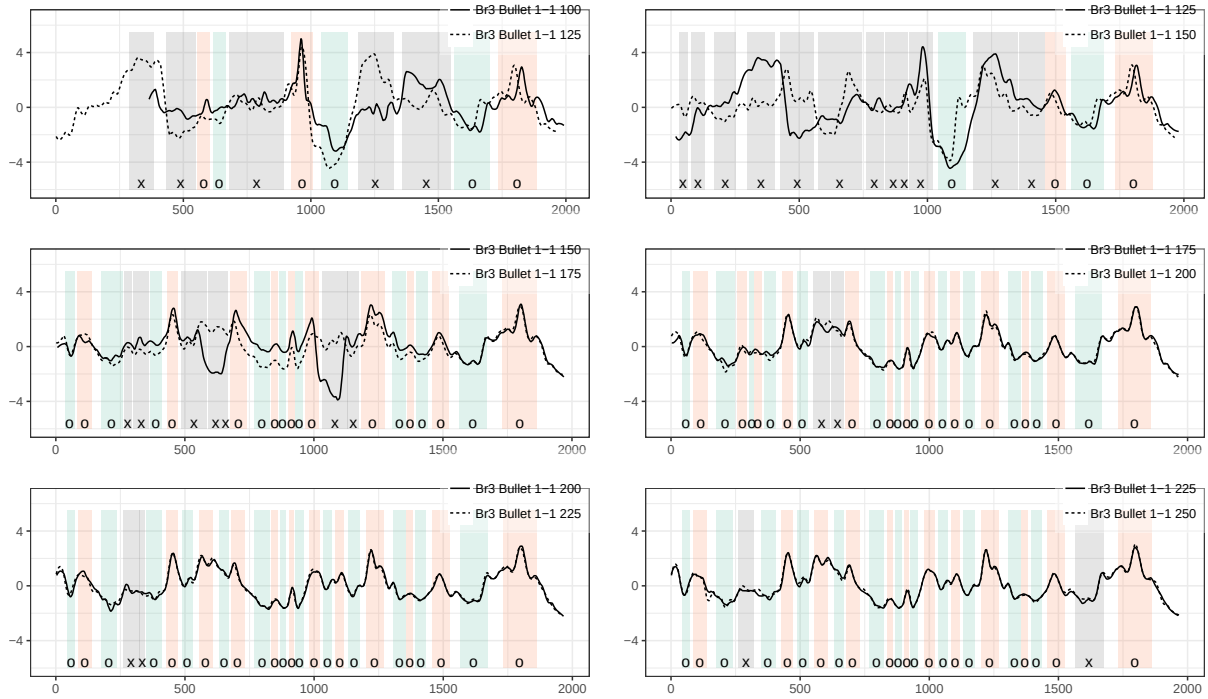


Figure 2.11: Signatures for barrel 3, bullet 1-1 extracted from varying heights. Initially, the match between signatures taken at heights $25\mu m$ apart is affected strongly by some break off at the bottom of the bullet. At a level of $175\mu m$ the bullet's signature stabilizes. For this land, matches should not be attempted at lower heights.

regression we selected a smoothing factor of 35, while for detecting the peaks/valleys of the loess residuals a smoothing factor of 25 seems more appropriate.

Figure 2.12 displays the peaks and valleys detected in the same signature at smoothing factors of 5, 25, and 45, respectively. The dark line corresponds to the smoothed values, while the grey line in the back shows the raw signature. The choice of smoothing factor is a classical decision of a bias/variance trade-off. It is immediately clear that a small smoothing factor like 5 is a poor choice. It results in a significant amount of noise in the data such that even just a point or two can skew the rolling average enough for a peak or valley to be detected. Given that striation widths are typically much larger, we are in effect muddying the waters by performing such minimal smoothing. Another consideration is that the smoothing should not fall below the resolution of the equipment at which the surface measurements are taken – so as to not introduce artifacts in the analysis.

A larger smoothing factor on the other hand (like 45), seems to be a more plausible option. Most of the peaks/valleys present which are detected by a smoothing factor of 25 are also detected at 45. However, some notable issues arise. Notice that the valley on the right hand side of the image is smoothed out, and thus not detected. On the left hand side, a double peak is detected - that might be a questionable decision - but there are several peaks in the middle, that are smoothed out, for example the peak at around $y = 750$. That is, in many cases, large windows are smoothing out some of the structure that we wish to see. Furthermore, it can be seen that the peaks/valleys are often shifted relative to their position in the original loess residuals, or in the smoothed data with smaller smoothing factors.

2.5 Evaluation

In order to get a better understanding of how the matching algorithm works in known matches and non-matches, we investigate its performance using the James Hamby study data. As a first step, we automatically assess the quality of each of the lands by checking that we can identify a stable region on each land. For this, we compute the cross-correlation of signatures

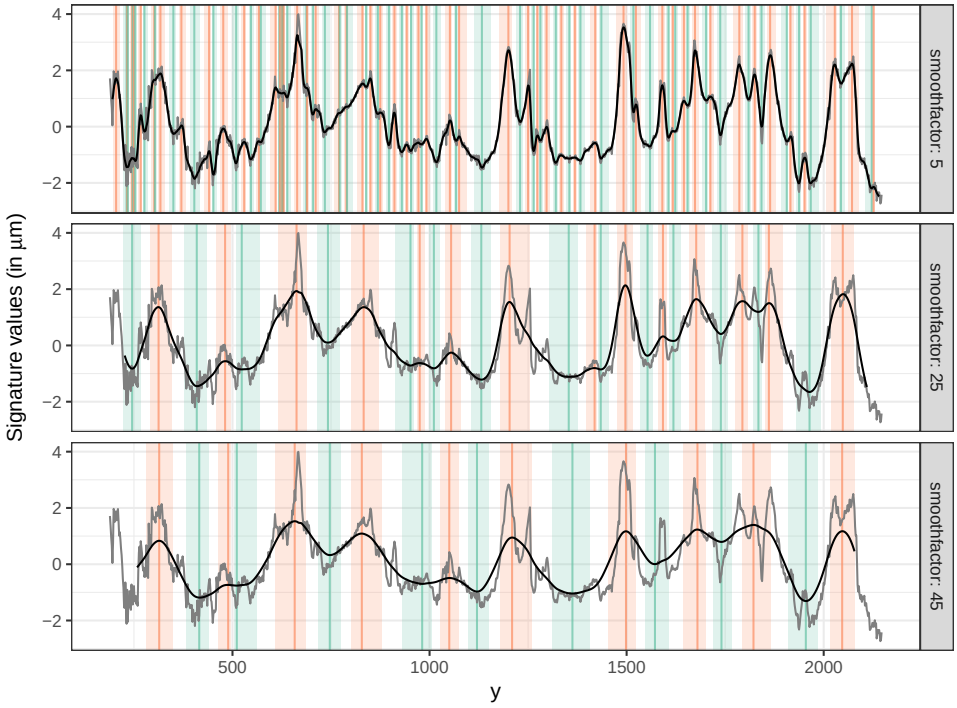
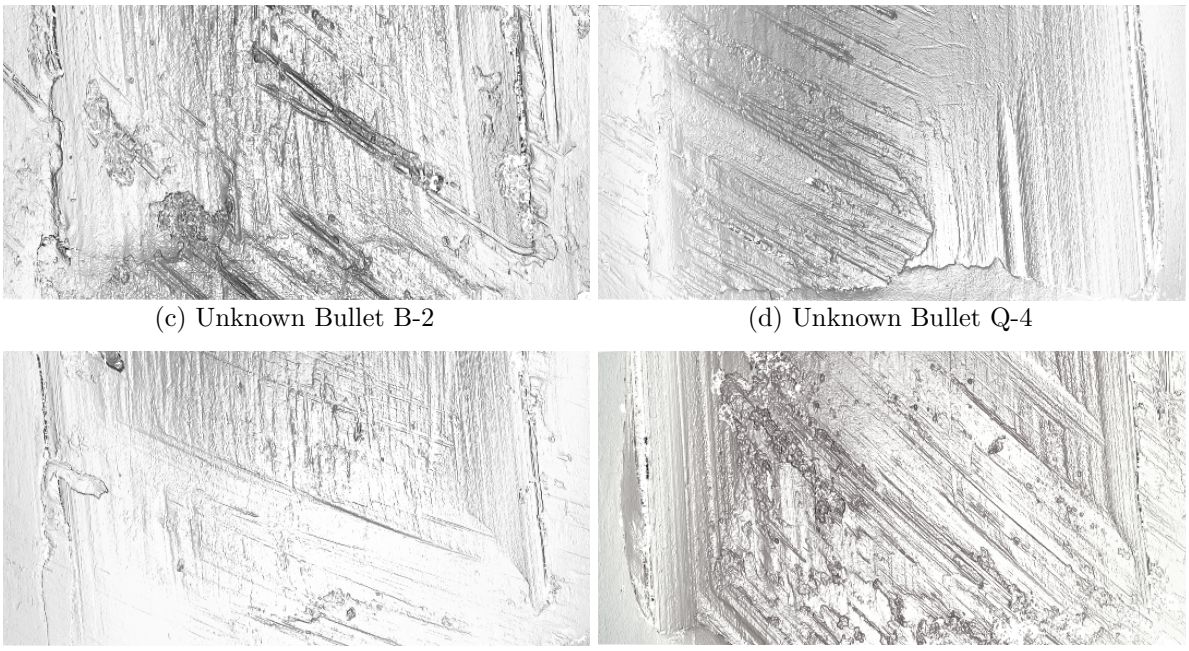


Figure 2.12: Peak/valley detection at smoothing factors of 5, 25, and 45, respectively. Note that a smoothing factor of 5 yields enough noise that many very minimal overlapping peaks and valleys are detected, while a smoothing factor of 45 might over-smooth and cause the peaks/valleys to either end disappear or shift horizontally from their original position in the signature.

extracted from heights $25\mu m$ apart. For a stable region, we require a minimum of 0.95 for the cross correlation. Four lands from different bullets are flagged as problematic in this respect. A visual inspection (see Figure 2.13) shows that each one of these lands has scratch marks across the surface, also known as 'tank rash'.

(a) Barrel 6 Bullet 2-1 (b) Barrel 9 Bullet 2-4



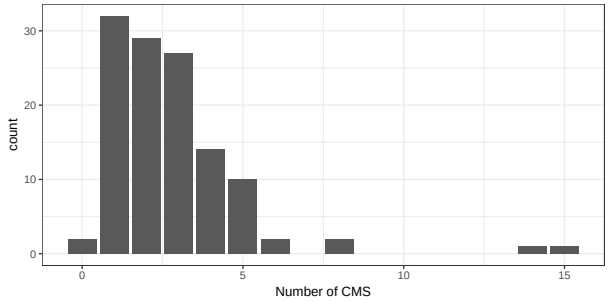
(c) Unknown Bullet B-2 (d) Unknown Bullet Q-4

Figure 2.13: Images of the four lands that got flagged during the quality assessment. All of them show scratch marks (tank rash) across the striation marks from the barrel. They are excluded from the remainder of the analysis.

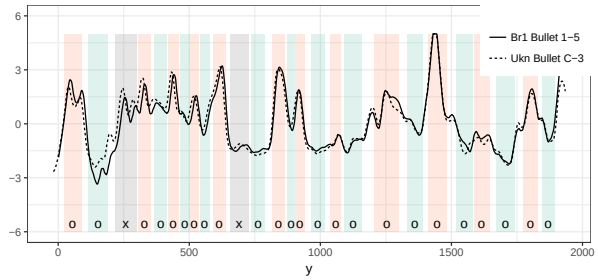
We exclude these four lands from further matching considerations and run all remaining lands from the unknown bullets against all remaining lands from known bullets for matches, i.e. we are comparing $15\times6-2 = 90-2 = 88$ lands from unknown bullets against $2\times10\times6-2 = 120-2 = 118$ lands from known bullets, yielding a total of $10,384$ land-to-land comparisons. Out of these comparisons, there are 172 known matches (KM), while the rest are known non-matches (KNM). Ideally, results look like the results in Figure 2.14: Figure 2.14a shows the distribution of the number of maximum consecutive matching striae between land C-3 and all 118 lands from known bullets. Two lands show a high CMS. These correspond to the known matches with C-3, shown in Figures 2.14b and 2.14c. Unfortunately, not all results are as clear cut. It might not be reasonable to assume that we can match all lands, but the idea is to try to maximize the

number of matches to get an overview of what we might be able to expect from an automated match.

(a) Maximal number of CMS between unknown bullet C-3 and all of the other 118 considered (known) lands. For two lands the number of maximum CMS is high.



(b) Overlaid signatures of C-3 and the land with the top matching CMS.



(c) Top 2 match with C-3 based on CMS.

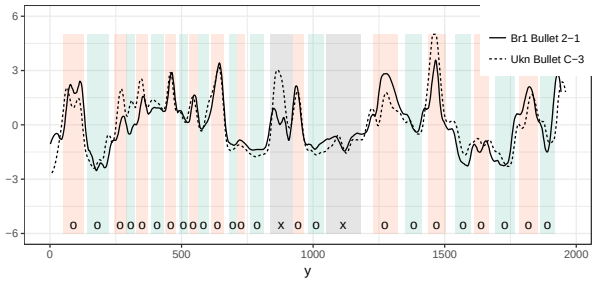


Figure 2.14: Showcase scenario when matching with CMS works very well. Unfortunately the matches are not always that convincing.

Figure 2.15 shows the strong connection between the maximal number of consecutive striae and matches in the Hamby study. All 42 pairs of lands with at least thirteen CMS in common are matches.

There are two things that should be noted at this point: the automated algorithm finds a relatively high number of CMS even for non-matches. On average, there are 2.31 maximal CMS between known non-matches (with a standard deviation of 1.4). Known matches share on average 8.49 maximal CMS, with a standard deviation of 5.65. While the probability for a match increases with the number of maximal CMS, a large number of maximal CMS by itself is not indicative of a match, as was previously pointed out by Miller (1998). Figure 2.16 shows a known mismatch between two lands that share twelve consecutively matched striae. Visually we can easily tell that these two lands do not match well.

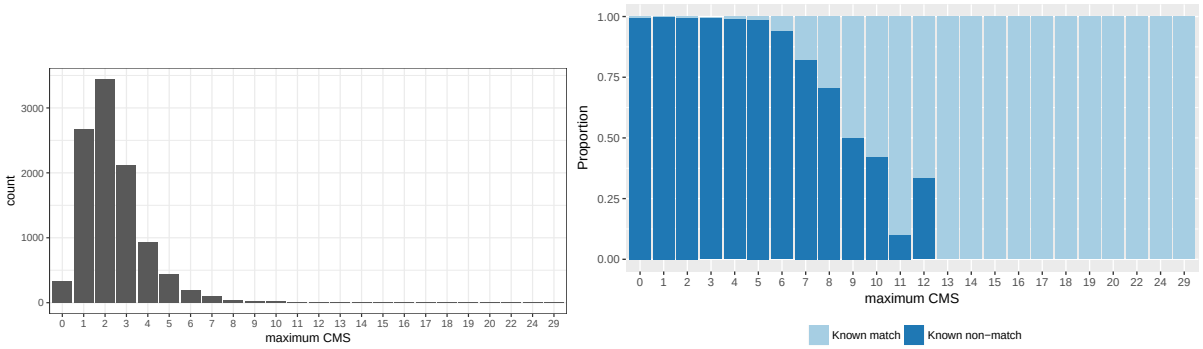


Figure 2.15: Distribution of maximal CMS (left). Conditional barchart (Hummel 1996) on the right: heights show probability of match/non-match given a specific CMS. All land-to-land comparisons with at least 13 CMS are matches.

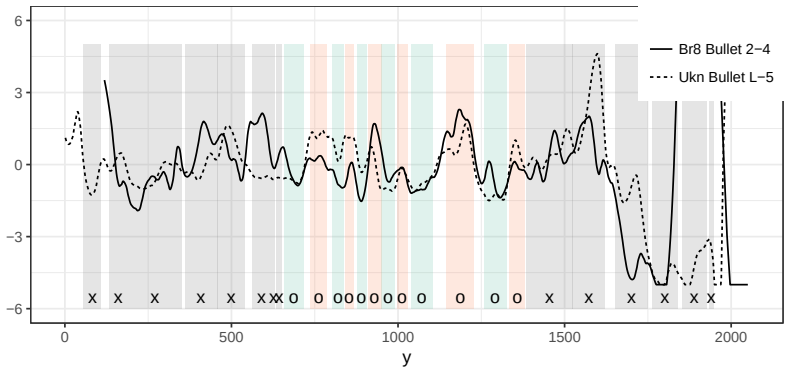


Figure 2.16: Known mismatch with a relatively large number of maximal consecutive matching striae (twelve) in the middle. The pattern in the middle does look surprisingly similar, however the outer ends of the signatures easily reveals this comparison as mismatch.

For smaller numbers of CMS, the percentage of false positives quickly increases. However, if we take other features of the image into account, we can increase the number of correct matches considerably: Figure 2.17 gives an overview of the densities of all of the features derived earlier, for known matches (KM) and known non-matches (KNM). The densities of almost all of the features show strong differences between matches and non matches. For example, a high amount of cross-correlation between two signatures is indicative of a match – in the Hamby study, only known matches have a cross-correlation of 0.75 or higher. There are 97 land-to-land comparisons with a cross-correlation that high.

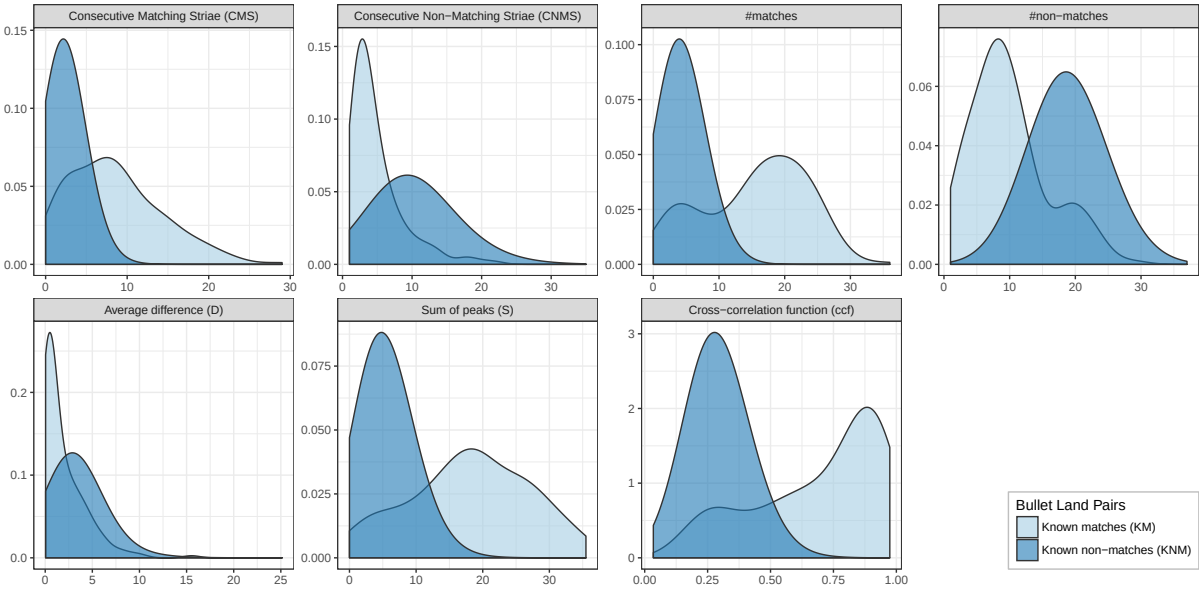


Figure 2.17: Overview of all the marginal densities for features described in Section 2.4.1. Shifts in the mode of the density functions between known matches and known non-matches indicate the variable's predictive power in distinguishing matches and non-matches. Predictive power is shown in more detail in Figure 2.18.

All of the features in Figure 2.17 show large, if not significant, differences between matches and non-matches. The predictive power of each one of these features is shown in the form of the Receiving Operating Characteristic (ROC) curves in Figure 2.18. The features are arranged in descending order according to the area under the curve (AUC). The dots mark the equal error rate, i.e. the location on the ROC curve, where false positive and false negative error rates are the same. The smaller the value, the better. We see that in this instance a low equal error rate (EER) goes hand in hand with high predictive power as measured in AUC. The feature with

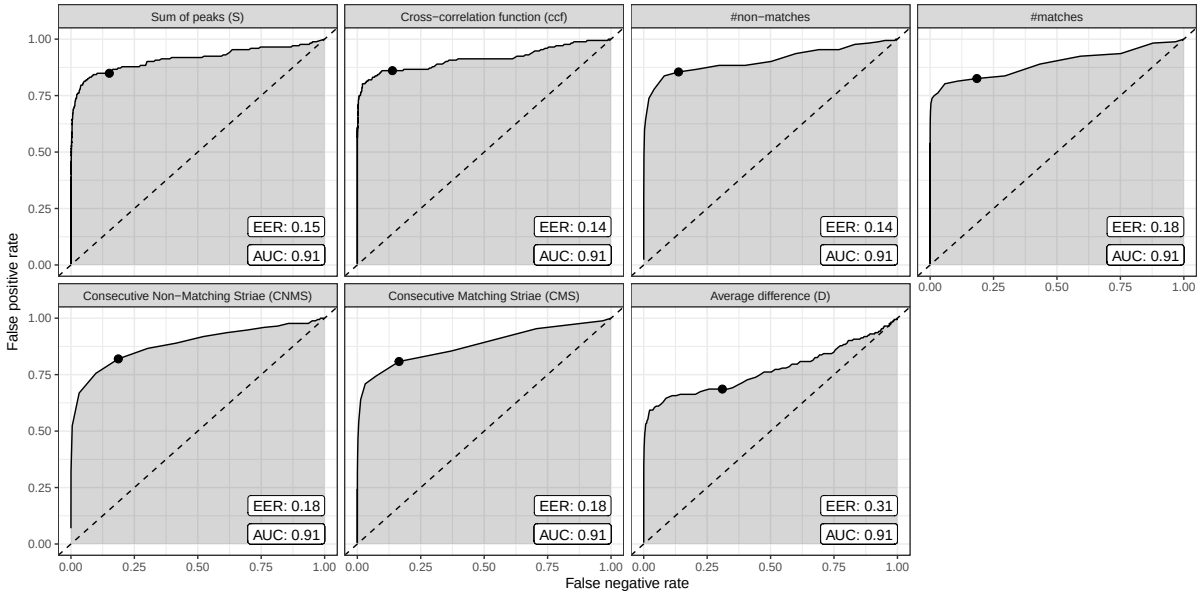


Figure 2.18: ROC curves for all of the features described in Section 2.4.1. Variables are sorted according to their area under the curve (AUC). The equal error rate (EER) is marked by a point on the ROC curve. Except for the distance D between signatures, all individual features derived from the surface measurements and the aligned striation marks are more predictive than the maximal CMS.

the highest individual predictive power is S, the sum of the average heights of two signatures at peaks and valleys. The maximal number of CMS is only in the seventh position here. The overall high AUC values indicate that we can successfully employ machine learning methods to distinguish matches from non-matches.

Using recursive partitioning, we fit a decision tree (Breiman et al. 1984, Therneau, Atkinson, and Ripley (2015), Milborrow (2015)) to predict matches between lands based on features derived from the image files. The resulting tree is shown in Figure 2.19. A total of 132 lands is being matched correctly. Interestingly, the number of consecutive matching striae does not feature in this evaluation. Instead of CMS, cross-correlation (ccf) between the signatures is very important in the matching process by the decision tree. Aside from cross correlation, the total number of matches is also included in the decision rule. Between cross-correlation and CMS, cross-correlation has higher predictive power. This does not contradict earlier findings emphasizing the value of CMS on visual assessments of bullet matches: in those papers, assessments were based on purely visual inspection of either actual bullets or 2D microscopic

images of bullets. Neither one of these methods allows for an assessment of cross-correlations. This is one of the benefits of switching to a digitized version of the images that preserves the 3D surface structure. The findings about the discriminating power of cross-correlation are consistent with the results of the study by Ma et al. (2004). However, in that study, the authors did not consider the number of matches and non-matches.

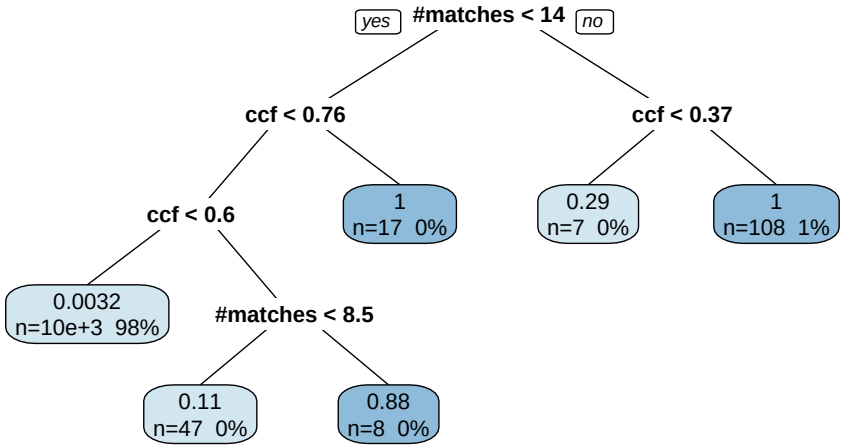


Figure 2.19: Decision tree of matching bullets based on recursive partitioning. The rectangular nodes are the leaves, giving a short summary consisting of the number of observations in the leaf (bottom left), the corresponding percentage of the total (bottom right). The number at the top shows the fraction of these observations that are a match. A 1 or a 0 therefore indicate a homogeneous (or perfect) node.

Another benefit of the digitized version of the images is that we can apply several hundred decision trees to combine in a random forest (Breiman 2001, Liaw and Wiener (2002)). For each of the trees in a random forest, only two thirds of the observations are used for fitting, while the remaining third is used to evaluate the tree's predictive power and accuracy, or its reverse, the error rate. Because errors are determined from the one third of held-back observations, this error rate is called the out-of bag (OOB) error. Figure 2.20 shows the cumulative out-of-bag error (OOB) rate for 300 trees.

After about 100 trees, the error rate of land-to-land comparisons stabilizes at 0.0039. This is a weighted average between false positive error rate of 0.0001 and an error rate of false negatives of 0.2267. This out-of-bag error rate is over-estimating the actual error in the Hamby study: here, the final random forest based on 300 trees is able to correctly predict all known matches and

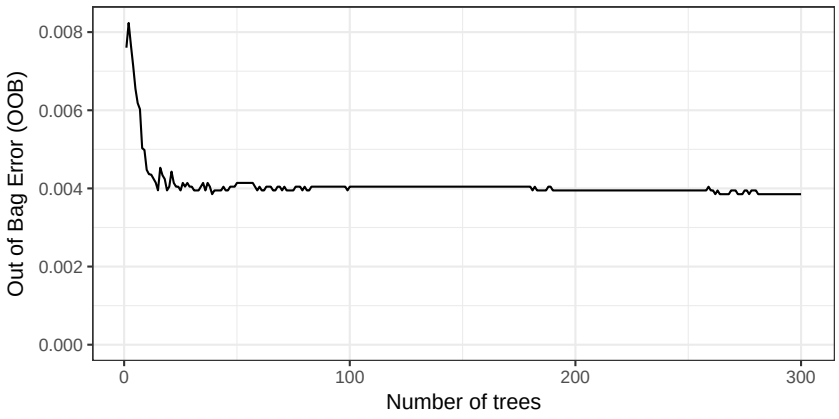


Figure 2.20: Cumulative out-of-bag error rate of a random forest fit to predict land-to-land matches from image features.

non-matches (see Figure 2.21). Note that this error rate is based on land-to-land comparisons and is expected to be much lower for bullet-to-bullet comparisons. In the case of the Hamby data, even a single tree results in an overall error rate of zero, if we require that a match of two bullets occurs when at least two of the bullet's lands are matched. This makes the errors in the automated approach smaller than the human error in the Hamby study. Out of the 507 participants who returned results, eight (out of $15 \times 507 = 7,605$) bullets were not matched conclusively, corresponding to a rate of 0.0011.

For the Hamby data, error rates based on bullet-to-bullet matches do not carry a lot of weight because of the small size of the study: fifteen unknown bullets are successfully matched to two pairs of ten bullets. Matching bullets can only be tested realistically in a much bigger experiment. Another thing to note about the random forest's error rates is that they are based on probability cutoffs of 0.5, i.e. whenever the predicted probability of a match exceeds 0.5, a match is declared. Basing this decision on a threshold fixed at 0.5 may not be the best approach. In practice, examiners are allowed a third option of 'inconclusive'. On a probability spectrum of outcomes we could therefore introduce an interval of 'inconclusive' results in the middle of the spectrum – which turns out to be unnecessary in the Hamby study, because, here, the results from the random forest are very clear cut. Figure 2.21 shows a comparison of the predicted probabilities of a match by the tree and the random forest. As expected, the random forest provides a more realistic estimate of the uncertainty in the classification.

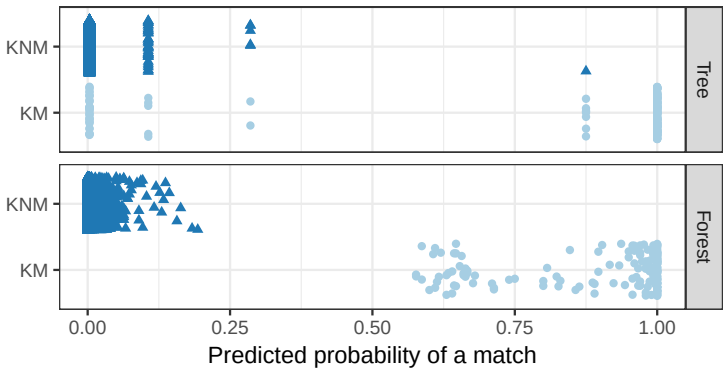


Figure 2.21: Prediction results from the tree and the forest. Using a cut-off probability of 0.5 the forest correctly predicts every single comparison. Compared to the tree, the forest's prediction probabilities are shrunk towards either end of the prediction range.

Besides resulting in a probabilistic quantification of matches, random forests also provide an assessment of the importance of each of the features derived from the bullets' 3D topological surface measurements. Figure 2.22 shows an overview of the importance of each variable measured as the mean decrease in the Gini index when the variable in question is included in a tree (for the exact values please refer to Section 2.8.5).

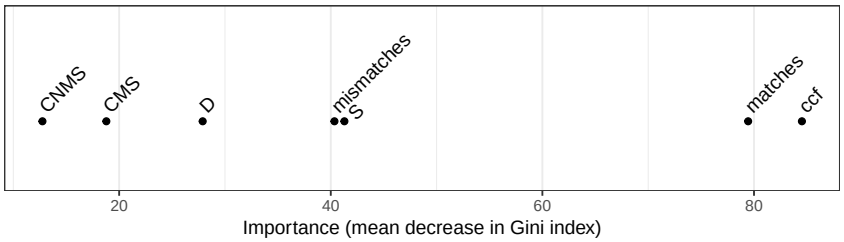


Figure 2.22: Importance of features in the random forest. Importance is measured in terms of mean decrease in Gini index when including the variable in a decision tree.

The variables with the most predictive power are cross-correlation and the overall number of matching extrema, followed by the total depth of joint striations S and total number of mismatches. CMS is found only in sixth place.

Besides including results from known matches against known non-matches, we can increase the number of comparisons in the Hamby study to include all possible land-to-land comparisons. This effectively doubles the number of data points available. Comparisons not previously

included in fitting the random forest can also be used as an additional source for assessing error rates. Results for this and a more detailed discussion can be found in Section 2.8.4.

2.6 Discussion

We present an algorithm which detects the most prominent but least relevant structure of a bullet from a firearms identification perspective, removes these features, and produces residuals which allow for the easy identification of markings. We have generalized this algorithm to align the residuals from two bullets to automatically determine whether they are indistinguishable. A random forest model provides a probabilistic assessment of the strength of a match, along with an ordering of the relevance of features. Matching bullets is clearly not a one-step process, but rather a sequence of data analysis tasks each deserving attention. As there is no scientific standard in place at this point in time, our intent is to explain an approach to addressing these tasks, while documenting all steps and providing all code so other researchers and forensic scientists can reproduce and expand on our findings.

The matching algorithm is sensitive to the parameter choices made. The heights at which signatures are extracted (currently 25μm apart) to evaluate stability, as well as the cross-correlation factor (currently 0.95) we set as a minimum threshold do affect the final outcome. Another parameter that must be selected is the amount of smoothing when identifying peaks and grooves (currently, a window of 23.4375μm is used, corresponding to a window of 7 values to the left and the right of an observation). We try to lay out in the paper the impact that each of the parameter choices has on the matching performance, but more research and better data are needed to define an optimized scenario.

The Hamby study serves as our evaluation 'database'. It consists of only 35 bullets – this is obviously not a particularly realistic scenario for an automatic matching procedure, but for now we are unaware of other databases containing bullets in the x3p format that we could add to our study.

The feasibility of creating a database of images that could be used to identify guns used in crimes was evaluated in a 2008 report (Committee to Assess the Feasibility, Accuracy and Technical Capability of a National Ballistics Database 2008) by the National Research Council. The committee investigated the scalability of NIBIN (National Integrated Ballistic Information Network), which uses proprietary matching algorithms provided by IBIS. The bottom line of the report was that in spite of the many technical and practical hurdles, solutions to all but one problem could be found. The problem that remained is that statistically, the quality of the matching algorithm (in this case, of breech-face marks and firing pin impressions) could not withstand a hugely increased number of records without overwhelming forensic examiners, who have to examine possible matches suggested by the system. The findings of the NRC report on imaging are based on two-dimensional greyscale images, which the committee argued were not reliable enough for distinguishing between fine marks. This finding coincides with the assessment by De Kinder, Tulleners, and Thiebaut (2004) based on the IBIS Heritage system. A further re-assessment by De Ceuster and Dujardin (2015) came to the same conclusions based on the EvoFinder system. The NRC report also found that results from 2D images can be improved when matches are based on 3D images. This is consistent with the importance of features found here: out of the top five features (see Figure 2.22), only the total number of matches and mismatches are available for a match based on 2D features.

By suggesting an automated algorithm that first removes class characteristics, such as the grooves and the curvature of the bullet to reveal the region of the land, then identifies peaks and valleys on this land, we reduce subjectivity and with it possible sources of bias. In particular, 'the concept of counting striations is subjective and based on experience' (Miller 1998). The steps outlined in this paper could also help explore other important forensic science problems. In particular, more general toolmark examination can benefit from the approach we discuss.

For a fair assessment of the performance of an algorithm, we need transparency. Our matching algorithm is open: the code is readily available in form of the R package bulletr (Hofmann and Hare 2016), and the code to produce this paper is available at http://www.github.com/erichare/imaging-paper. To understand whether an automated approach along the lines of

the one we propose can accurately identify sets of bullets with undistinguishable markings, it will be necessary to assemble a much larger database that includes a wide range of ammunition types, degrees of damage, gun makes, etc. We are unaware of the existence of any such database. In addition to serving as a realistic testbed for the performance of the automated matching algorithm, such a database would also permit testing the underlying, as of yet untested, assumptions of uniqueness and reproducibility of the markings left by a gun on bullets.

2.7 Acknowledgment

Thanks to David Baldwin for pointing us to the NIST database and doing a Firearms 101 for us. Thanks to the men and women behind the software R (R Core Team 2016), and the authors of the R packages knitr (Xie 2015) and ggplot2 (Wickham 2009).

2.8 Appendix

2.8.1 Cylindrical Fit

Figure 2.23 shows the profile of surface measurements of bullet 1-5 at a fixed height. The smooth line on top is a circle, with estimated radius and center. The details of this fit are given below:

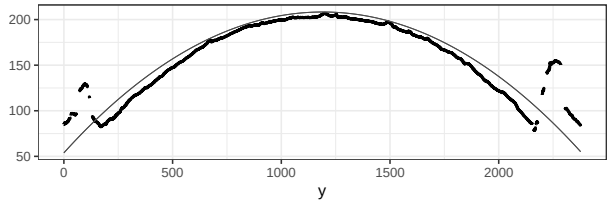


Figure 2.23: Side profile of the surface measurements (in μm) of a bullet land at a fixed height of x. Note that the global features dominate any deviations, corresponding to the individual characteristics of striation marks.

Assume that n data points are given in the form of data tuples (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) that are (approximately) located on a circle. We want to estimate the location of the center and radius of the best fitting circle using a least squares approach.

We minimize the following expression:

$$D = \sum_{i=1}^{n} \left(r^2 - (x_i - a)^2 - (y_i - b)^2 \right)^2, \tag{2.1}$$

by differentiating D with respect to r, a, and b: let us assume that x_i and y_i are centered (i.e. $\sum x_i = \sum_i y_i = 0$). Note, if they are not, make a note of the current means, subtract them now and add them to $(\hat{a}, \hat{b})$ at the end.

The derivate of D with respect to r is:

$$\begin{aligned} \frac{d}{dr} D &= 2 \sum_i \left(r^2 - (x_i - a)^2 - (y_i - b)^2 \right) 2r = \\ &= 4r \left(nr^2 - \sum_i (x_i - a)^2 - \sum_i (y_i - b)^2 \right). \end{aligned}$$

At the minimum:

$$\frac{d}{dr} D = 0 \overset{r \neq 0}{\iff} nr^2 = \sum_i (x_i - a)^2 + \sum_i (y_i - b)^2. \tag{2.2}$$

The derivative of D with respect to a is:

$$\begin{aligned} \frac{d}{da} D &= 2 \sum_i \left(r^2 - (x_i - a)^2 - (y_i - b)^2 \right) 2(x_i - a) = \\ &= -4 \left[a \cdot nr^2 + \sum_i (x_i - a)^3 + \sum_i (x_i - a)(y_i - b)^2 \right]. \end{aligned}$$

Using (2.2) for nr^2 in the equation above we get:

$$\begin{aligned}
\frac{d}{da}D &= -4\left[\sum_i a(x_i-a)^2 + \sum_i a(y_i-b)^2 + \right. \\
&\qquad \left. \sum_i (x_i-a)^3 + \sum_i (x_i-a)(y_i-b)^2\right] = \\
&= -4\left[\sum_i (x_i-a)^2(a+x_i-a) + \right. \\
&\qquad \left. \sum_i (x_i-a+a)(y_i-b)^2\right] = \\
&\qquad\qquad\qquad\qquad\qquad\qquad \textstyle\sum_i x_i = 0 \\
&= -4\left[\sum_i (x_i-a)^2 x_i + \sum_i x_i(y_i-b)^2\right] \quad \textstyle\sum_i y_i = 0 \\
&= -4\left[\sum_i x_i^3 + \sum_i x_i y_i^2 - 2as_{xx} - 2bs_{xy}\right],
\end{aligned}$$

where $s_{xx} = \sum_i x_i^2, s_{xy} = \sum_i x_i y_i$ and $s_{yy} = \sum_i y_i^2$.

Likewise, we get for the derivative of D with respect to b:

$$\frac{d}{db}D = -4\left[\sum_i y_i^3 + \sum_i x_i^2 y_i - 2as_{xy} - 2bs_{yy}\right].$$

To find the minimum we therefore get a system of two linear equations in a and b:

$$\begin{aligned}
2s_{xx}a + 2s_{xy}b &= c_1 \qquad \text{with } c_1 = \sum_i x_i^3 + x_i y_i^2 \\
2s_{xy}a + 2s_{yy}b &= c_2 \qquad \text{with } c_2 = \sum_i x_i^2 y_i + y_i^3.
\end{aligned}$$

The solution to the system is:

$$\begin{aligned}\hat{a} &= \frac{c_1 s_{yy} - c_2 s_{xy}}{2s_{xx}s_{yy} - 2s_{xy}^2},\\ \hat{b} &= \frac{c_2 s_{xx} - c_1 s_{xy}}{2s_{xx}s_{yy} - 2s_{xy}^2}, \text{ and}\\ \hat{r^2} &= \frac{1}{n}s_{xx} + \frac{1}{n}s_{yy} + \hat{a}^2 + \hat{b}^2.\end{aligned}$$

The scatterplot in Figure 2.24 shows the residuals of such a fit. In this instance, the radius is estimated as $\hat{r} = 4666.49\mu m = 4.67mm$ and the land covers about 29.5 degrees. Both of these estimates are consistent with a 9 mm bullet fired by a Ruger P-85. The residuals are dominated, as expected, by the grooves, which show up as large positive residuals. For a profile at height $x = 100\mu m$ there is a residual circular structure that does not show up for all signatures.

(a) Residual structure at height $x = 1.5625\mu m$ (bottom of the bullet).

(b) Residual structure at height $x = 100.00\mu m$

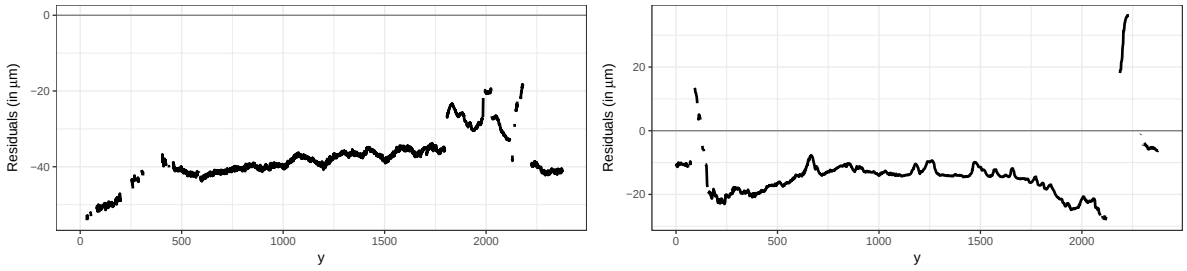


Figure 2.24: Residual structure of circular fits at two different cross sections. Both residual plots show systematic structures, indicating that a circular fit is not entirely appropriate.

A single cylinder as a fit is unlikely to be a particularly good fit, because there seem to be quite massive deformations in the vertical direction. Even when we fit a circle at each distinct height of the bullet, as in Figure 2.25, this does not address all of these issues. While the wider circumference at the base of the bullet can be resolved by individual circular fits, the systematic residual structure in Figure 2.24b stays the same.

2.8.2 Cross-Correlation at Multiple Heights

Figure 2.26 shows a sequence of signatures for bullet 1-5 (barrel 1) that are taken at heights $50\mu m$ apart, between $150\mu m$ and $400\mu m$. These are compared to the signature at a height

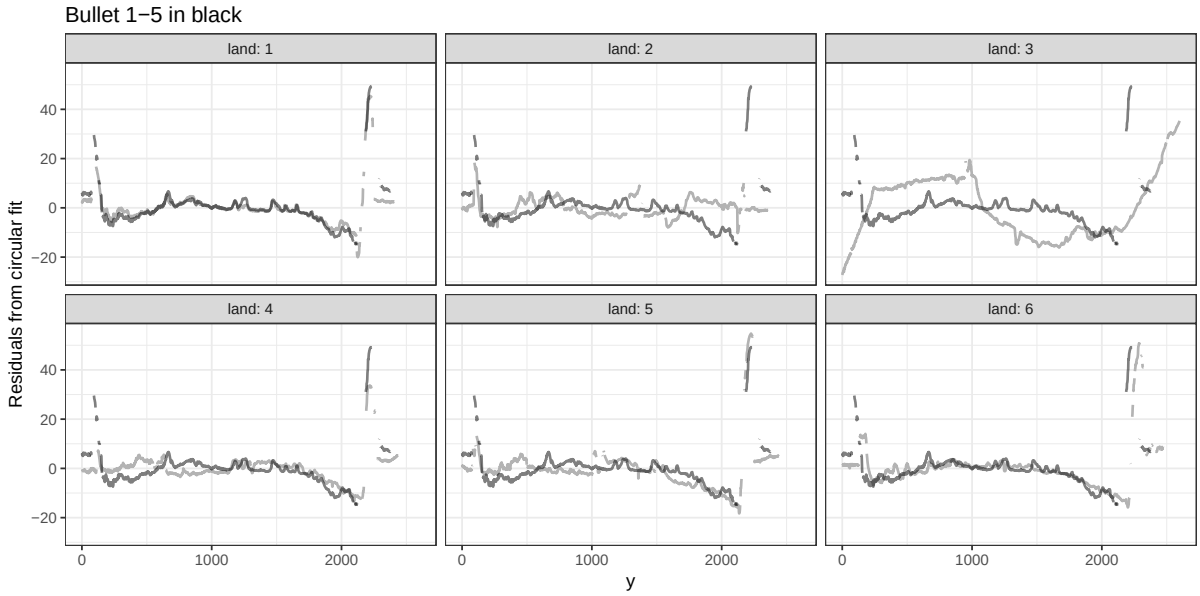


Figure 2.25: Circular fit to the signature of each land of bullet 2, with signature from bullet 1-5 overlaid. The signature of bullet 1-5 matches best with bullet 2-1.

of $100\mu m$. Initially, this comparison constitutes an almost perfect match between the two signatures. However, the match quickly deteriorates with increasing distance between the heights at which signatures are extracted. Only if signatures are from heights within $150\mu m$ do we get a good visual match even when we know that the same bullet surface is being used. Given that we have to expect some minor variation in the same height values due to (manual) alignments in microscopes, we should take height values into account in the automatic matching routine by evaluating matches at several heights.

2.8.3 Signature Intensities

Figure 2.27 shows an overview of the signatures at different heights on a single bullet.

At larger heights individual characteristics become less distinctive, making true matches to other bullets harder. The pattern of decreasing peaks and valleys is generally true for bullet lands, as can be seen in Figure 2.28.

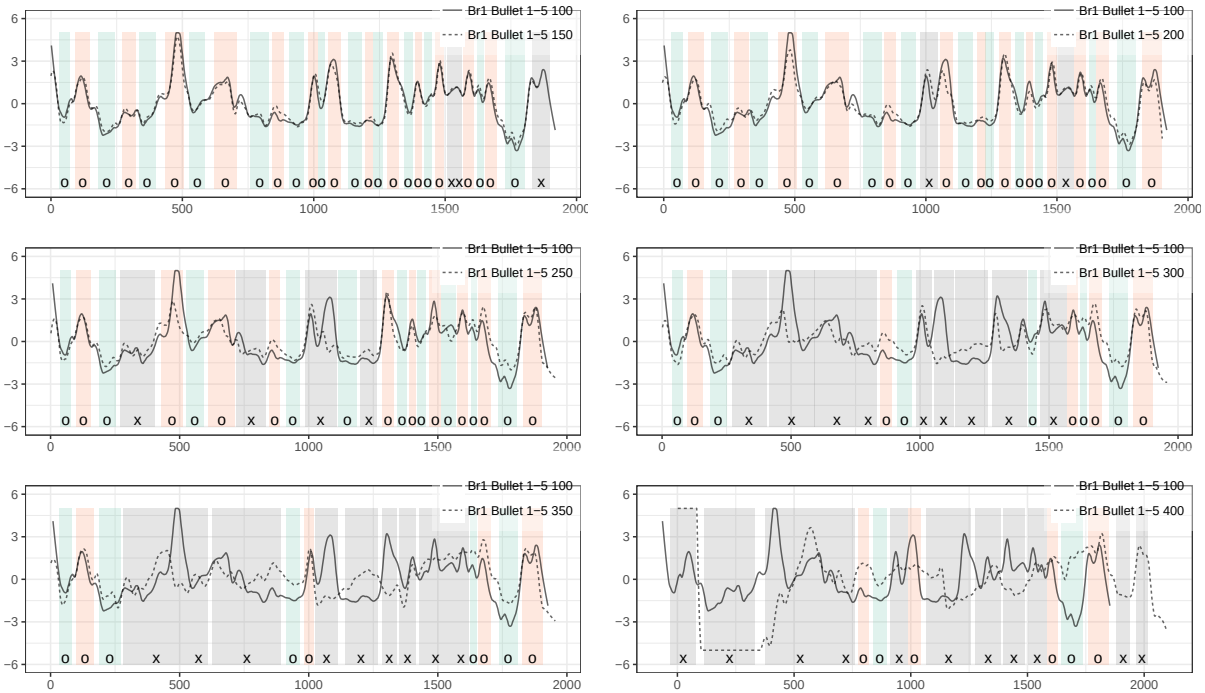


Figure 2.26: Overview of the variations in the signatures at different heights. The signature extracted at $x = 100\mu m$ is compared to signatures at every $50\mu m$. With every step away from the original height, the number of differences between the signatures increases; the number of maximum CMS decreases from initially 22 to four or fewer at a height of $x = 300\mu m$ and above.

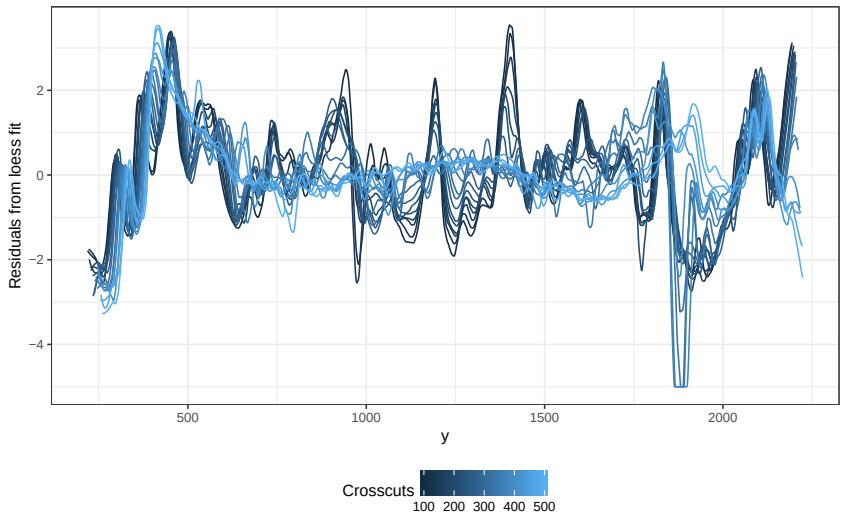


Figure 2.27: Signatures of the same bullet at different heights. With increasing height, peaks and valleys are less pronounced, resulting in a smaller standard deviation.

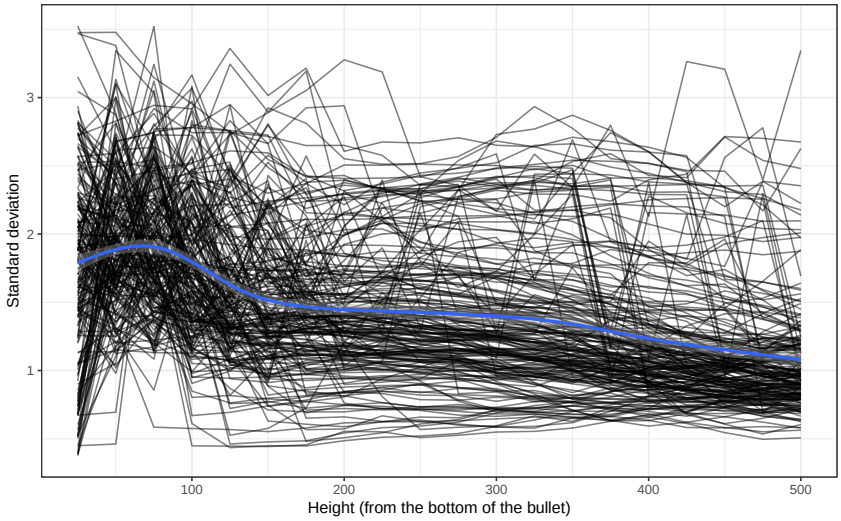


Figure 2.28: Standard deviation reduces as height increase.

Figure 2.28 shows that the amount of standard deviation of a signature decreases on average for all bullet lands at larger heights. This makes standard deviation of a signature one measure to quantify the extent to which a signature is expressed. For identifying matches we should therefore use the lowest height to extract a bullet's signature once a stable surface region is detected. This is in accordance with current standard practice (AFTE Criteria for Identification Committee 1992).

2.8.4 Complete Evaluation of the Hamby Study

One way to expand the use of the James Hamby study is to not only match all of the unknown bullet lands against the known bullet lands, but to compare every land against every other land. This effectively doubles the number of comparisons from 10,384 pairwise comparisons of usable bullet lands to 21,115 $[= (118 + 88) \cdot 205/2]$ comparisons by adding another 10,731 bullet land comparisons made up of known-to-known and unknown-to-unkown comparisons.

When we predict the new 10,731 comparisons using the random forest based on the previously considered 10,384 known-unknown comparisons, we encounter 18 false negatives and 9 false positives, corresponding to an actual false error rate of 0.19 and a false positive rate of 0.00085, which is close to the random forest's estimated OOB error rates of 0.226744 and 0.000098.

However, if we use all of the available comparisons to fit another random forest of 300 trees, the defacto error rates for false positives and false negatives are again at 0. The estimated OOB error rates are 0.00024 for the false positive rate and 0.22180 for the false negative rate. The false positive rate is therefore virtually unchanged, while we see a slight improvement in the false negative rate for an overall OOB error rate of 0.3%, i.e. an increase to twice the number of comparisons leads to a decrease of 25% of the estimated error rate. This is yet another argument in favor of a larger database for training algorithms.

Figures 2.29 and 2.30 give an overview of all the signatures from bullet lands in the Hamby study aligned by barrel. Three to five bullets were fired from each barrel. The figures give us both some insight into how well signatures match and how consistent individual characteristics are impregnated on bullets fired from each of the barrels. Signatures for some lands match remarkably well – such as land 5 from barrel 1, whereas all lands from barrel 5 show some variability both in the location and depths of peaks and valleys.

2.8.5 Table of Feature Importance

Two random forests were calculated for the Hamby study. For the first random forest only comparisons of bullet lands from known bullets and unknown bullets were used. The second random forest is based on a full comparison of every land with every other land, increasing the number of comparisons from originally 10,384 (10,212 known non-matches and 172 known matches) by another 10,931 comparisons (10,637 known non-matches and 94 known matches). Random forests allow an assessment of variable importance (also called feature importance) as the mean decrease in Gini index when including each variable. Table 2.1 shows the results for feature importance for both of these random forests. Importance 1 refers to the smaller subset, Importance 2 is the feature importance derived from the random forest based on all pairwise comparisons.

Figure 2.29: Overview of aligned signatures for all bullet lands for barrels 1 to 5 of the Hamby study.

Figure 2.30: Overview of aligned signatures for all bullet lands for barrels 5 to 10 of the Hamby study.

Table 2.1: Table of features derived from bullet image ordered by importance in predicting matches. Importance is measured in terms of mean decrease in Gini index when including the variable in a decision tree. Averages (and standard deviations) for known matches (KM) and known non-matches (KNM) are shown in the last four columns.

	Variable	Importance 1	Importance 2	KM	(sd)	KNM	(sd).1
1	ccf	87.0	134.6	0.7	(0.25)	0.3	(0.10)
2	#matches	81.9	128.3	15.5	(7.91)	4.3	(2.51)
3	S	46.4	53.6	18.3	(8.95)	5.5	(3.41)
4	#non-matches	35.9	62.7	9.8	(5.80)	18.8	(3.92)
5	D	26.1	45.7	1.9	(2.32)	3.3	(1.94)
6	CMS	15.9	25.1	8.5	(5.65)	2.3	(1.40)
7	CNMS	13.3	20.9	4.8	(4.10)	10.2	(4.35)

CHAPTER 3. ALGORITHMIC APPROACHES TO BULLET MATCHING WITH AN EMPHASIS ON THE DEGRADED LAND CASE

A paper to be submitted to the **Journal of Forensic Science**.

Eric Hare, Heike Hofmann, Alicia Carriquiry

Abstract

Bullet matching is a process used to determine whether two bullets have been fired from the same gun barrel. While traditionally a manual process performed by trained forensic examiners, recent work has been done to add statistical validity and objectivity to the procedure. In this paper, we build upon the algorithms explored in Automatic Matching of Bullet Lands (Hare, Hofmann, and Carriquiry 2016) by formalizing and defining a set of features, computed on pairs of bullet lands, which can be used in machine learning models to assess the probability of a match. We then use these features to perform an analysis of the two Hamby (Hamby, Brundage, and Thorpe 2009) bullet sets (Set 252 and Set 44), to assess the presence of microscope operator effects in scanning. We also take some first steps to addressing the issue of degraded bullet lands, and provide a range of degradation at which the matching algorithm still performs well. Finally, we discuss generalizing land to land comparisons to full bullet comparisons as would be used for this procedure in a criminal justice situation.

3.1 Background

Intense scrutiny has been focused on the process of bullet matching in recent years (e.g., Giannelli 2011). Bullet matching, the process of determining whether two bullets were fired from the same gun barrel, has traditionally been performed without meaningful determination of error rates or statistical assessments of uncertainty (National Research Council 2009). There have been certain steps towards mathematical and statistical approaches to bullet matching, including defining CMS, the Consecutively Matching Striae (Biasotti 1959), and defining a cutoff of six to separate matches from non-matches. But rigorous assessments of the applicability of such cutoffs have not to this point been described (Advisors on Science and Technology 2016).

Recently, work has been done to address these well-known shortcomings. On firing pin impressions and breech faces, Riva and Champod (2014) have described an automated algorithm using 3D images. Other examples of work in this and related areas include Petraco and Chan (2012), Chu et al. (2011), Vorburger et al. (2011), and others. In our approach to this problem, Automatic Matching of Bullet Lands, we used the Hamby 252 set (Hamby, Brundage, and Thorpe 2009) to train and develop a random forest in order to provide a matching probability for two bullet lands (Hare, Hofmann, and Carriquiry 2016). While the algorithm had a very strong performance on this set, some limitations were immediately clear. For instance, performance was assessed only on this single set of 35 bullets fired from a consecutively manufactured set of only ten gun barrels. Each of these bullets was part of controlled study, and the full lands were available for matching. While there were some data quality issues, this was still a near ideal test case for the algorithm.

Real world applications of bullet matching often involve the recovery of fragments of bullets from the crime scene. Traditional features used in forensic examination work well for a full land, but there has been less investigation into their performance in the case of a fragmented land. For example, the CMS is naturally limited by the portion of the land that can be recovered.

In this paper, we take steps to address these and other concerns. Specifically, we begin by reviewing features from the literature, computed on pairs of bullet lands, and providing some of our own features. We show how these have been standardized to account for the portion of the land that is recovered. Once these have been standardized, we tackle two issues that were previously unaddressed in Automatic Matching of Bullet Lands. The first is the effect of the operator of the microscope on the resulting data and algorithm performance, and the second is the effect of the amount of degradation. Finally, we take a few first steps towards generalizing a matching algorithm based on land-to-land comparisons, to one based on bullet-to-bullet comparisons, as would be done in a real world application of these ideas.

3.2 Feature Standardization

To start, we introduce a standardized version of each feature used in the matching routine. These features are computed on *aligned pairs of bullet lands* rather than individual lands. This allows us to, for instance, compute the number of matching striae. Table 3.2 and Table 3.3 provide an example of six land-to-land comparisons and the derived features for these comparisons. The two `profile_id` columns identify a particular land from the two Hamby datasets, and the remaining columns are the derived features for these comparisons.

Table 3.2: A sample of six land-to-land comparisons, and derived features for these comparisons.

profile1_id	profile2_id	ccf	rough_cor	lag	D	sd_D
49	540	0.2689	0.1091	-0.1734	0.0026	0.0043
49	1032	0.3878	0.3582	0.1938	0.0017	0.0030
49	1540	0.2410	0.1320	0.2781	0.0023	0.0039
49	2044	0.2892	0.1136	0.0828	0.0023	0.0038
49	2583	0.1769	0.0214	-0.0563	0.0026	0.0041
49	3074	0.2715	0.1254	0.0938	0.0020	0.0033

The definitions of these features have been generalized to account for the possibility of handling degraded bullet conditions, where only fragments of lands can be recovered. The definitions of each feature are given below, where $f(t)$ represents the height values of the first profile, and $g(t)$ the height values of the second:

Table 3.3: The remaining derived features for the previous six land-to-land comparisons.

signature_length	overlap	matches	mismatches	cms	non_cms	sum_peaks
1.9328	0.9103	1.1368	10.2251	0.5684	3.7182	2.0038
1.9359	0.9193	5.0571	6.5593	3.3714	2.3426	6.5264
1.9359	0.8757	1.7696	10.3662	1.1797	5.8592	2.0192
1.9172	0.9829	1.5920	7.9441	0.5307	3.4756	1.8139
1.8234	0.9692	1.1317	8.8479	0.5659	4.9155	0.8555
1.9172	0.9690	2.6913	10.7645	1.0765	3.4251	3.5272

- **ccf** (%) is the maximum value of the Cross-Correlation function evaluated at the optimal alignment. The Cross-Correlation function is defined as $C(\tau) = \int_{-\infty}^{\infty} f(t)g(t+\tau)dt$ where τ represents the the lag of the second signature (Vorburger et al. 2011).
- **rough_cor** (%) is a new feature representing the correlation between the two signatures after performing a second LOESS smoothing stage and subtracting the result from the original signatures. This attempts to model the roughness of the surface after removing the waviness.
- **lag** (mm) Is the optimal lag for the ccf value.
- **D** (mm) is the Euclidean vertical distance between each height value of the aligned signatures. This is defined as $D^2 = \frac{1}{\#t}\sum_t [f(t) - g(t)]^2$. This is a measure of the total variation between two functions (Clarkson and Adams 1933).
- **sd_D** (mm) provides the standard deviation of the values of D from above.
- **signature_length** (mm) is the overall length of the smallest of the two aligned signatures.
- **overlap** (%) provides the percentage of the two signatures that overlap after the alignment stage.
- **matches** (per mm) is the number of matching peaks/valleys (striae) per millimeter of the overlapping portion of the aligned signatures.
- **mismatches** (per mm) is the number of mismatching peaks/valleys (striae) per millimeter of the overlapping portion of the aligned signatures.
- **cms** (per mm) is the number of consecutively matching peaks/valleys (striae) per millimeter of the overlapping portion of the aligned signatures (Biasotti 1959, Chu et al. (2013)).

- **non_cms** (per mm) is the number of consecutive mismatching peaks/valleys (striae) per millimeter of the overlapping portion of the aligned signatures.
- **sum_peaks** (per mm) is the the sum of the average heights of matched striae.

The features that are provided on the per millimeter level are intended to support the degraded land case, as discussed. Note that the computation differs slightly depending on the feature. For example, to standardize the number of matches, the raw number of matching striae are taken, and this is divided by the length of the overlapping region of the two lands (`overlap` from above). In most cases, the overlapping region will be very close to the length of the smaller signature. But depending on the alignment, this may not always be true. This ensures that we don't punish a particular cross-comparison for having a smaller region in which matches could occur. On the other hand, the number of mismatches is divided by the total length of the two aligned signatures, since mismatched striae can occur even in the non-overlapping region of the two signatures.

The `rough_cor` or Roughness Correlation is derived by performing a second smoothing step, and subtracting the result from the original signatures. This creates a new signature which eliminates some of the overall structure, allowing global deformations to have less of an influence on the model output. Where the roughness correlation is most useful is in a scenario like Figure 3.31. This figure shows the alignment of profile 40977 with 47600. The top panel shows the smoothed signatures. The middle panel overlays a LOESS fit to the average of the two signatures. Finally, to derive the roughness correlation, this LOESS is subtracted from the original signature to create a new set of roughness residuals, which are then given in the bottom panel. Note that these two profiles do not match, yet the ccf is 0.7724. The roughness correlation (-0.0324) correctly indicates the lack of matching. The roughness correlation acts as a check against false positives which can arise when there are significant deformations in the overall structure, as in the case with both these profiles.

In a typical comparison between two profiles, such as in Figure 3.32, the roughness correlation does not meaningfully impact the matching probability given the presence of the ccf in the

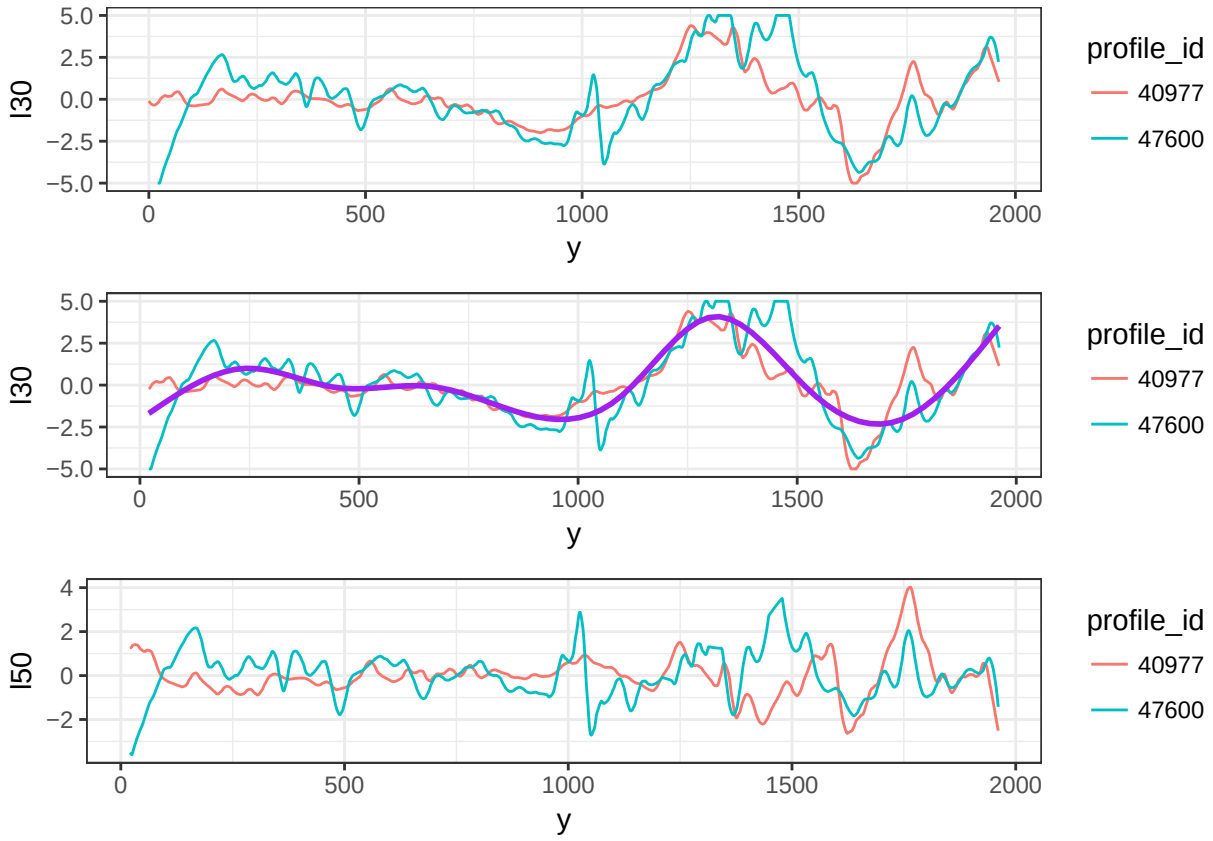


Figure 3.31: Alignment of profile 40977 with 47600. The top panel shows the smoothed signatures. The middle panel overlays a LOESS fit to the average of the two signatures. Finally, to derive the roughness correlation, this LOESS is subtracted from the original signature to create a new set of roughness residuals, which are then given in the bottom panel. Note that these two profiles do not match, yet the ccf is 0.7724. The roughness correlation (-0.0324) correctly indicates the lack of matching.

model. In this figure, we see the alignment of profile 8752 with profile 136676. In this case, the waviness or the deformation pattern in the signatures is more minor, and hence the resulting roughness signature resembles the original signature more closely. These profiles match, and both ccf (0.6891) and rough_cor (0.7980) provide values indicative of matching.

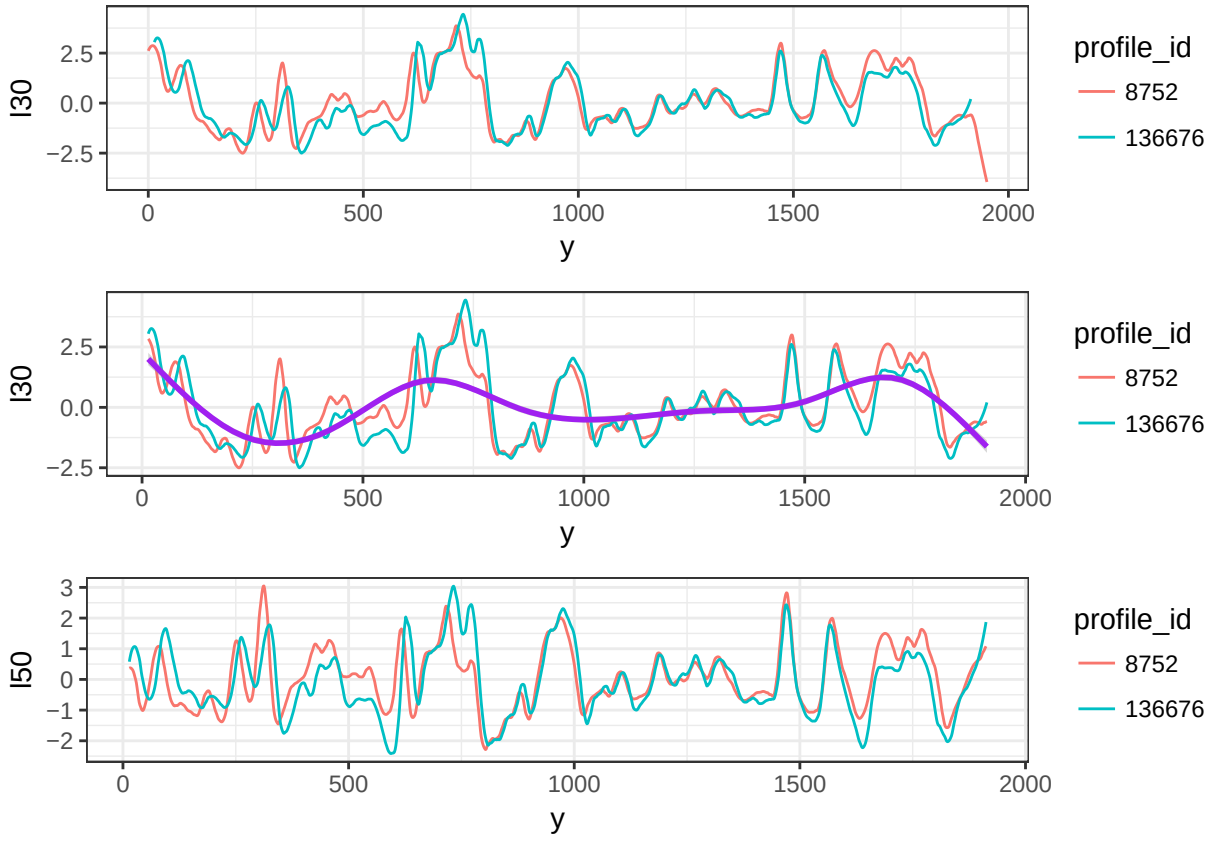


Figure 3.32: Alignment of profile 8752 with profile 136676. In this case, the waviness or the deformation pattern in the signatures is more minor, and hence the resulting roughness signature resembles the original signature more closely. These profiles match, and both ccf (0.6891) and rough_cor (0.7980) provide values indicative of matching.

We can observe the distributions of both CCF and the Roughness Correlation side by side, differentiating between known matches and known non-matches. Figure 3.33 shows this. It can be seen that the separation of the two groups along both variables is quite strong, but many known-matches have low values for each. The distributions of known non-matches, on the other hand, is relatively symmetric and normal.

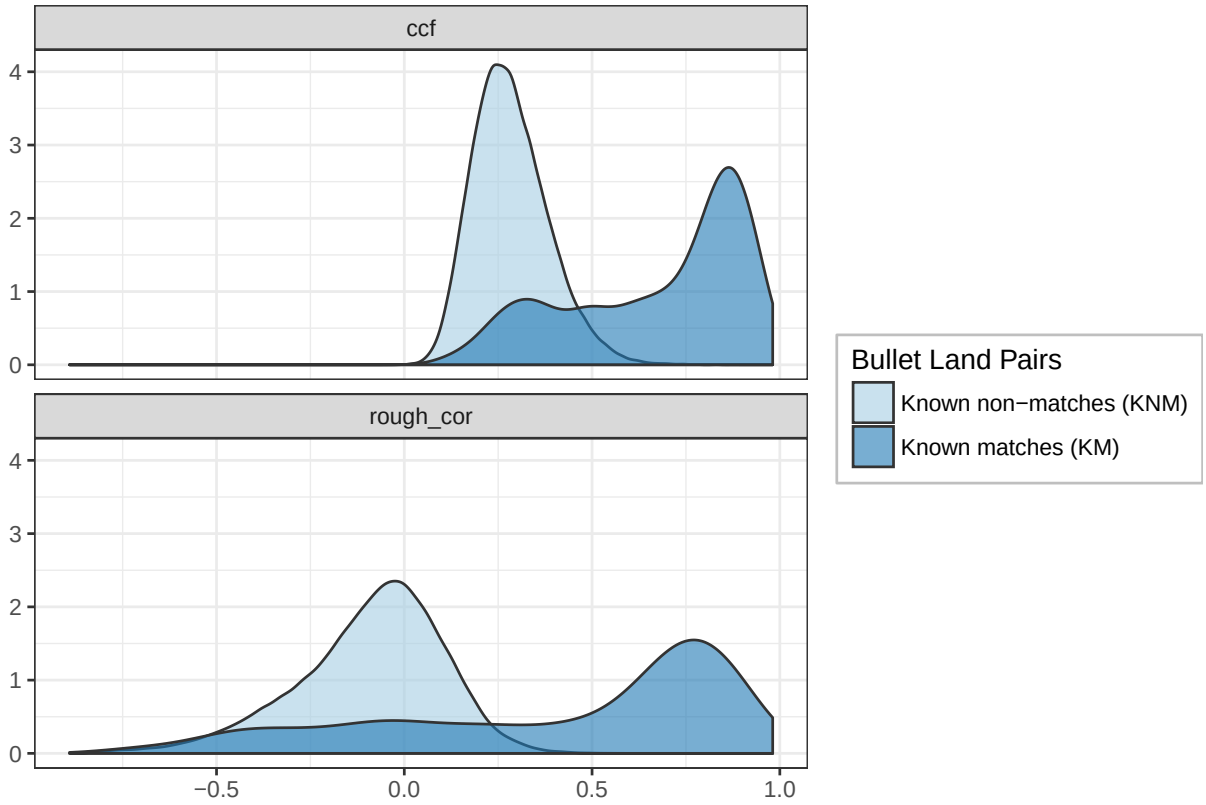


Figure 3.33: Distributions of the Roughness Correlation compared to the CCF for known matches and known non-matches

3.3 Model Training

Using these features, we can train a randomForest (Liaw and Wiener 2002) model which attempts to predict whether two lands match given the values of these features. We first join information on the study the data originated from, along with information on whether they are matches. The three studies currently contained in the database are Hamby (Set 252), Hamby (Set 44), and Cary. For purposes of the remainder of this analysis, we will exclude the Cary bullets from consideration. Because this was a study specifically designed to assess the persistence of striation markings over a series of fires from the same barrel, every Cary bullet is a known match to every other Cary bullet, although early firings do not produce the same markings that later firings do. Hence, we will consider Hamby (Set 252) and Hamby (Set 44) only.

We can now train the forest using the previously defined features. We split the data into an 80% training, 20% testing framework to assess its out of sample performance, using the `caret` package (Jed Wing et al. 2016). Table 3.4 provides the results as a confusion matrix on the test set, averaged over ten independent random forests trained on ten random data subsets. It can be seen that false positives are exceedingly rare, but false negatives occur more frequently (approximately 65 false negative land to land comparisons on the test set, compared with an average of less than four false positives).

Table 3.4: The average confusion matrix for the 10 random forests. It can be seen that false positives are exceedingly rare, but false negatives occur more frequently.

Result	Count
False Negative	65.3
False Positive	3.8
True Negative	16363.1
True Positive	171.8

These results suggest that our algorithm is a bit too conservative in predicting a match when in fact the bullets were fired from the same gun barrel. We can break down the confusion matrix depending on the study that each of the two lands originated from. Table 3.5 provides the average confusion matrix for the 10 random forests, broken down by study. It can be seen that Hamby252 to Hamby252 comparisons exhibit the fewest errors, while Hamby252 to Hamby44 comparisons exhibit the most on average. This intuitively makes some sense given the presence of potential scanner operator effects, which we will address further in this section.

Table 3.5: The average confusion matrix for the 10 random forests, broken down by study. It can be seen that Hamby252 to Hamby252 comparisons exhibit the fewest errors, while Hamby252 to Hamby44 comparisons exhibit the most on average.

Study	False Negative	False Positive	True Negative	True Positive
Hamby252_Hamby252	0.29%	0.01%	98.77%	0.93%
Hamby252_Hamby44	0.47%	0.02%	98.35%	1.16%
Hamby44_Hamby44	0.34%	0.04%	98.72%	0.9%

3.4 Feature Robustness

As a first stage to assessing feature robustness, we can produce parallel coordinate plots of the various features based on true positives, true negatives, false positives, and false negatives. Figures 3.34 and 3.35 provide this. The means of the true positive and the false negative groups are shown, respectively, in the two figures. False positives tend to occur with anomalously high values of `sum_peaks`, `matches`, and `ccf`, while false negatives tend to resemble very closely the feature distribution of the true negatives, with a slightly higher average `ccf`.

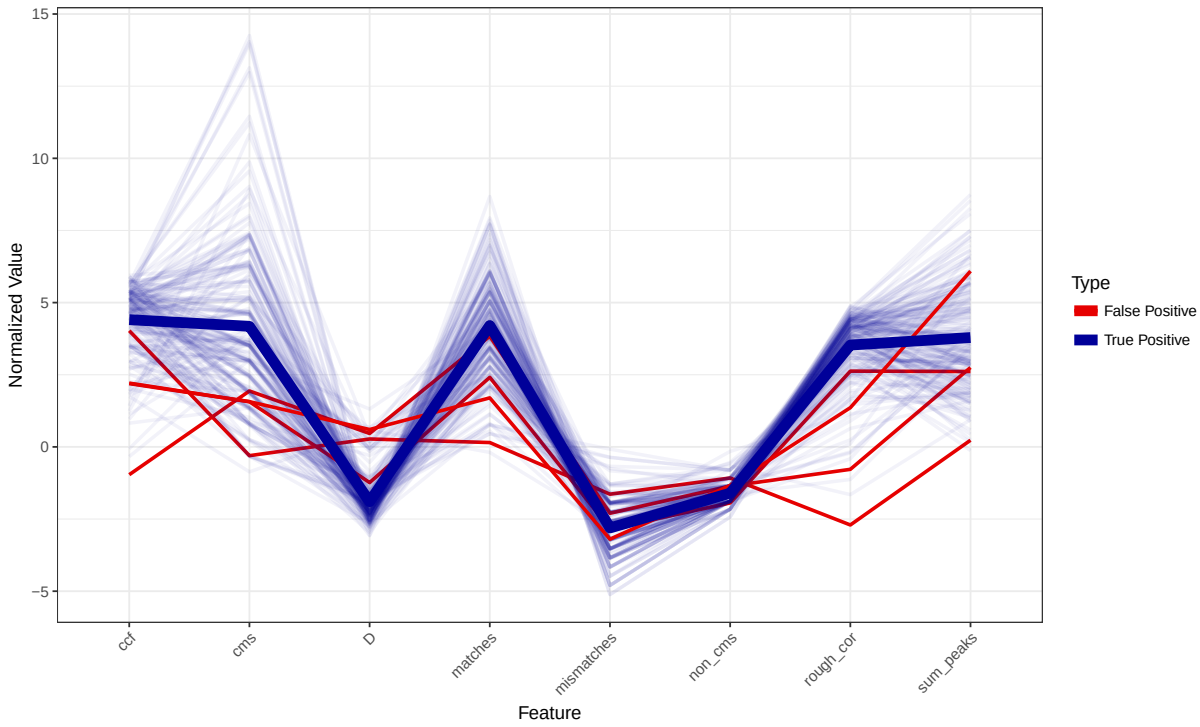


Figure 3.34: Parallel coordinate plot of the features based on the random forest confusion matrix for true and false positives.

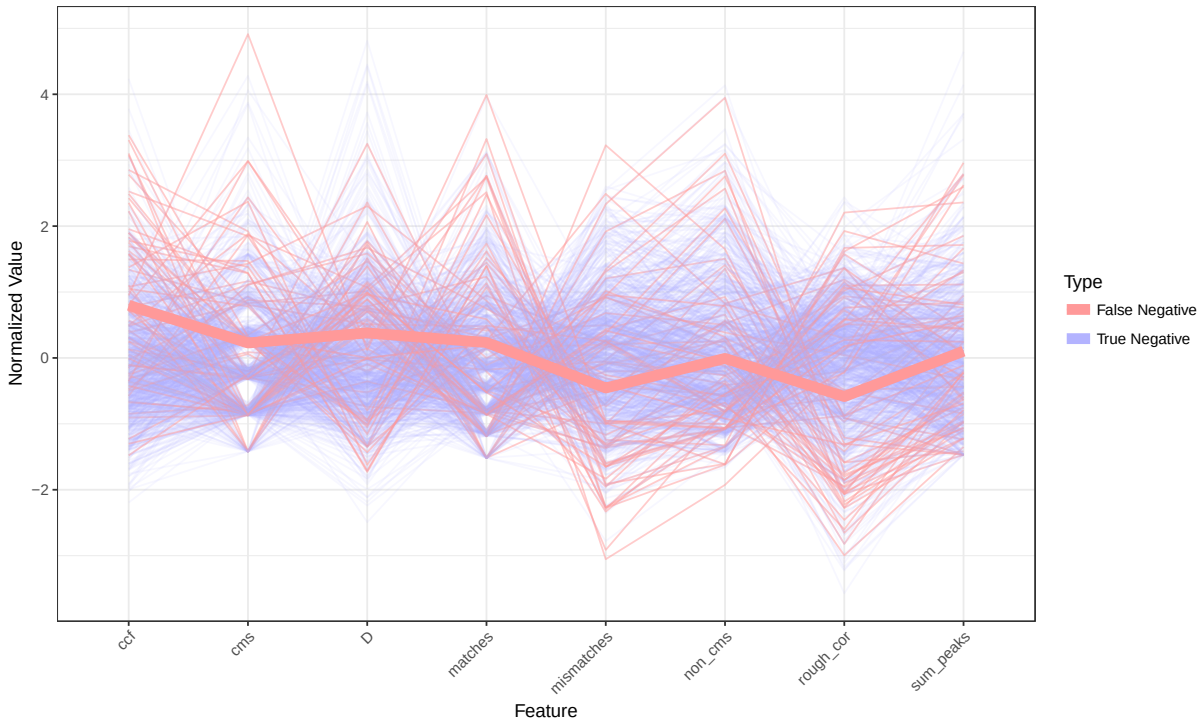


Figure 3.35: Parallel coordinate plot of the features based on the random forest confusion matrix for true and false negatives.

3.4.1 Operator Effects

We can attempt to quantify the effect of the study on the matching probability by fitting a new random forest which attempts to predict the study based on the derived features. Ideally, if the assumption of independence between lands holds across different operators, this forest should have poor performance - The set of derived features should be relatively consistent among known matches and known non-matches regardless of the study since the Hamby data in both sets originated from the same gun barrels.

Table 3.6 provides the confusion matrix, with column proportions, for the random forest with study as the response. It can be seen that while overall the random forest performs poorly, as hoped, comparisons between Hamby252 bullets are more distinguishable from other comparisons.

Table 3.6: Confusion Matrix (Column Proportions) for the random forest with study as the response. It can be seen that while overall the random forest performs poorly, as hoped, comparisons between Hamby252 bullets is more distinguishable from other comparisons.

Prediction \ Actual	Hamby252_Hamby252	Hamby252_Hamby44	Hamby44_Hamby44
Hamby252_Hamby252	28.5%	10.23%	3.24%
Hamby252_Hamby44	70.57%	83.27%	79.78%
Hamby44_Hamby44	0.93%	6.51%	16.98%

Figure 3.36 give the distributions of the features defined above, faceted by whether the lands are known to be fired from the same gun barrel, across different study to study comparisons. The distributions among the known non-matches seem relatively consistent across study based on visual inspection. On the other hand, among known matches, Hamby252 to Hamby252 comparisons exhibit more pronounced features, including a higher average ccf, higher number of matches, and higher value of sum_peaks.

Though visual inspection clearly shows differences, we can more formally assess the differences in the distributions formally with a Kolmogrov-Smirnov test. Table 3.9 gives the results of

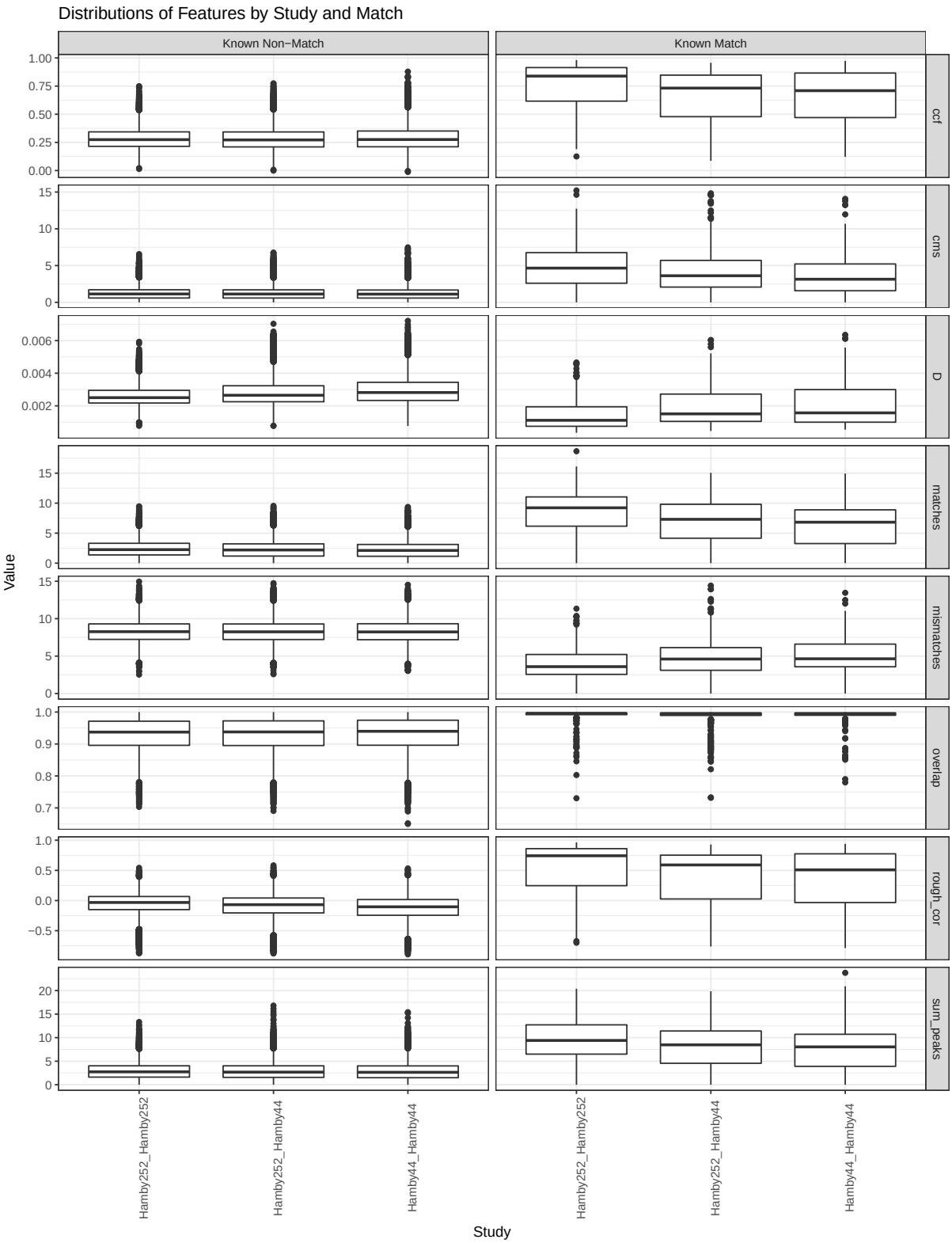


Figure 3.36: Distribution of the features, facetted by match, for different study to study comparisons of lands.

pairwise tests, for each feature, between different set comparisons, and between known matches compared with known non-matches. Although the tests are significant, looking at the raw values of the D statistic suggest that the largest effect sizes do in fact occur in comparisons with two Hamby252 lands, as the visual inspection of the boxplots also suggested.

These results strongly suggest the need for controlling for more effects when performing the analysis. Specifically, microscope operator effects resulting in variations in scan quality and scan parameters seem to play a role in the utlimate performance of the algorithm. Land to land comparisons from Hamby252 consistently result in more pronounced expression of features among known matches, and therefore result in a better ultimate accuracy in the random forest. Rigorous procedures to ensure scan quality and consistency across operators should be put in place to minimize the effect of the study and ensure the assumption of land to land independence is satisfied.

Another way to demostrate the study/operator effect is by observing the distribution of our algorithm's ideal cross section by study. Figure 3.37 gives the distributions of the ideal cross sections by study. It can be seen that the Hamby44 ideal cross sections are much more likely to be close to the base of the bullet compared to Hamby252.

Indeed, another Kolmogorov-Smirnov test also confirms a significant difference in the distributions of these values ($D = 0.6239, p < 0.0001$). Because the difference is significant, it strongly suggests that the operator effect in the bullet scanning procedure must be taken into account in order to assume pairwise independence.

3.4.2 Degraded Lands

We now turn our attention to matching degraded bullet lands, in which only fragments of the land can be recovered. Because the NIST database currently contains only full bullet lands, this will involve performing a simulation and making some simplifying assumptions. We will simulate various levels of degradation from the left, right, and middle of the land. We will use

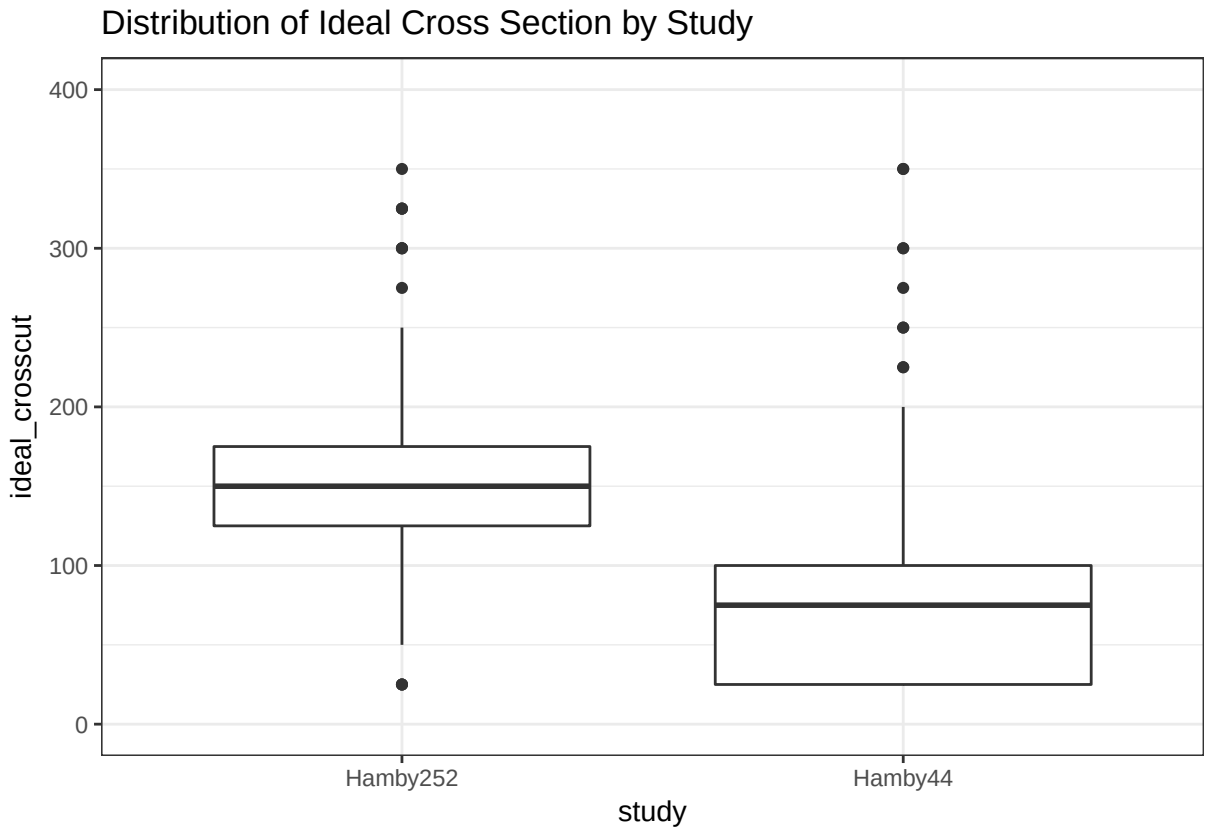


Figure 3.37: Distributions of the ideal cross sections by study. It can be seen that the Hamby44 ideal cross sections are much more likely to be close to the base of the bullet compared to Hamby252.

land proportions ranging from 100% to 25%, where the percentage represents the proportion of the land that was recovered. For example, a left-fixed 75% scenario implies that the left hand portion of the land was recovered, and the 25% rightmost portion was lost. We will do this by subsetting the signatures. Note that this is a bit of a simplification because the signatures themselves are somewhat dependent on the data that is missing because of the properties of the LOESS smoother.

Figure 3.38 gives the sensitivity (true positive rate) and specificity (true negative rate) of the random forest predictions for given levels of degradation. It can be seen that the sensitivity drops a bit until 50% land proportion, and then rises. This occurs because the algorithm begins producing more positive predictions in general, likely as a result of the ccf being arbitrarily higher for known non-matches due to the small signature. On the other hand, the specificity drops dramatically for left, middle, and right fixed degraded lands for land proportions of less than 50%. For a more in-depth exploration into the matching probabilities, Figure 3.45 provides histograms of the matching probability facetted by the degradation level and known match versus known non-match. The matching probabilities suffer compared with the full lands in all cases. The jump seems to be most noticeable beginning at about 25% degradation (75% land proportion), and the algorithm struggles beyond 50%.

Figure 3.39 gives the feature expression for known matches, as a function of the land proportion. It is immediately visible that the feature expression becomes quite variable when only a small fraction of the land is recovered, such as 25%. For instance, `sum_peaks` and the `cms` both drop, while `D` rises. Interestingly, some of the features are better expressed for the middle-fixed case. Overall, the feature expression remains relatively consistent as the land degrades up to a bit under 50% land proportion.

To come full circle, we will now attempt to match a particular land which exhibits bad tank rash. Figure 3.40 provides an image of the surface of this land. Due to the tank rash, it was originally excluded from matching consideration. However, it can be seen that approximately

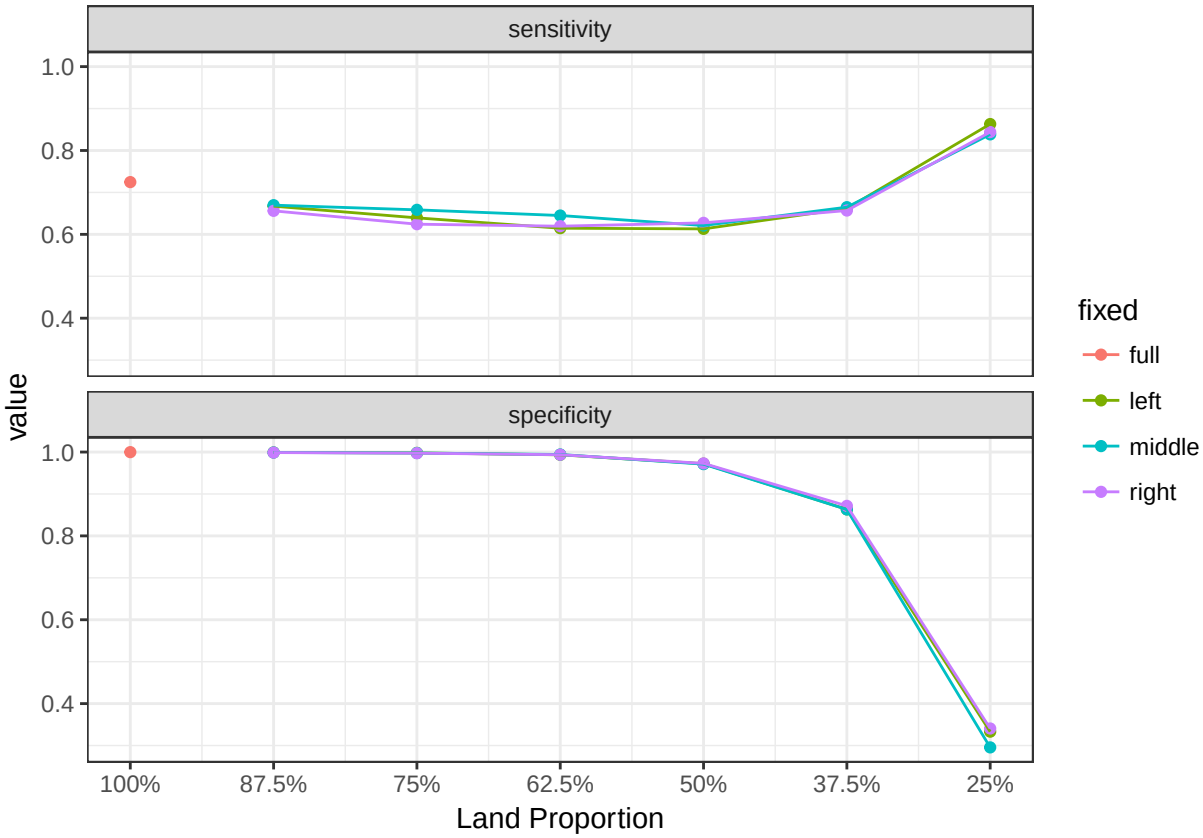


Figure 3.38: Sensitivity and specificity of the random forest for given levels of degradation. It can be seen that both metrics decline as a function of the land proportion, except for the sensitivity, which rises for very low levels of the land proportion due to an increase in the amount of positive predictions.

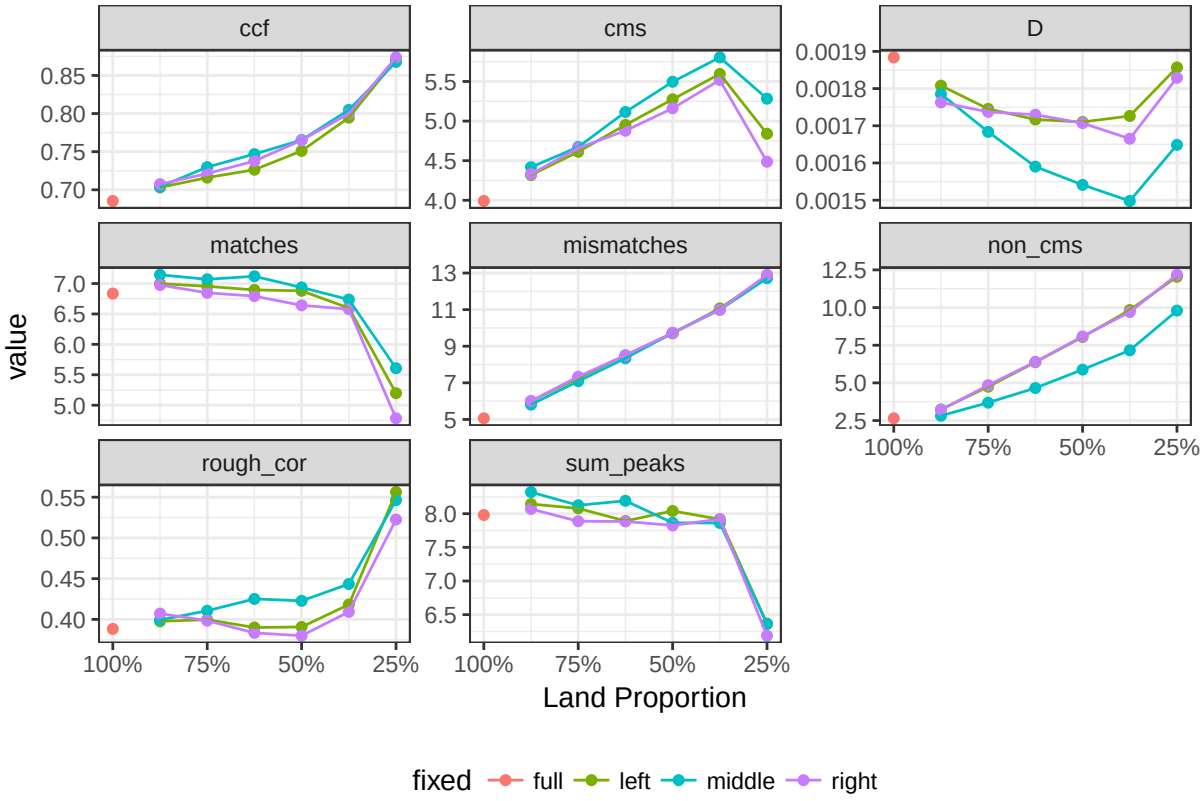


Figure 3.39: Feature expression for known matches, as a function of land proportion. It can be seen that when we fix the middle portion of the bullet land, the features tend to be better expressed.

half of the bullet remains relatively unaffected. We will extract a signature of the unaffected half and attempt to match this signature to its full known match.

Figure 3.40: Land 4 of Bullet 2, from Barrel 9 of Hamby Set 252. It can be seen that this particular land exhibits some major tank rash on the right half.

Table 3.7 provides the features derived, after extracting only the first 50% of the Hamby Barrel 9 Bullet 2, 4th land (and hence, simulating a left-fixed 50% degraded scenario), compared with a feature comparison between both full lands (and hence, including the tank rash striae). The features are derived in a comparison with its known match, the full Bullet 1 third land fired from Barrel 9. The features, including the ccf and the matches, are expressed enough to (barely) indicate a match in the case of the degraded bullet. Using the pre-trained random forest, the predicted matching probability is 52%. This is encouraging in that attempting to match the full bullet land, by comparison, yields a matching probability of 0.0067%. This is due to the relatively higher values of the ccf, cms, and matches for the degraded comparison, and suggests the feature standardization is working as intended.

Table 3.7: Features extracted for a comparison of the full Hamby Barrel 9 Bullet 1 Land 3, with a left-fixed 50 percent degraded portion of Hamby Barrel 9 Bullet 2 Land 4. These two lands are known matches, and indeed the random forest does predict a match.

Feature	Degraded Land	Full Land
ccf	0.6004	0.4442
rough_cor	0.3671	0.1633
D	0.0018	0.0023
overlap	0.9968	0.9968
matches	10.2236	5.6275
mismatches	7.5949	5.0713
cms	9.2013	4.6043
non_cms	6.5823	2.5357
sum_peaks	12.0020	6.3148
matchprob	0.5200	0.0067

3.5 From Lands to Bullets

One other area deserving further exploration is generalizing these algorithms for matching full bullets rather than indvidual lands, as would be done in a criminal justice application. One such approach is to recognize that (at least for the Hamby bullets) there should be six matching pairs of lands for any two bullets that were fired from the same gun barrel. Therefore, for each pair of bullets, we can extract the six highest matching probabilities and average them. If we do so, we obtain a clear separation as seen in Figure 3.46. No known-matches have a score below 50%, while all known non-matches have a score below 10%.

We can improve on this approach by exploiting the rotation of the bullet to compute a score. Under the assumption of land to land independence, we can define the probability of two bullets matching (M) as one minus the probability that the two bullets do not match (NM). Exploiting the idea that given two bullets do not match, none of the individual lands match, we can write the matching probability as the probability that at least one land pair in the matrix matches. Specifically,

$$P(M) = 1 - P(NM) \tag{3.3}$$

$$= 1 - (P(NM1) \times P(NM2) \times ... \times P(NM6)) \tag{3.4}$$

$$= 1 - ((1 - P(M1)) \times (1 - P(M2)) \times ... \times (1 - P(M6))) \tag{3.5}$$

Where M is the event that two bullets match, NM is the event that two bullets do not match, $M1$, $M2$, ..., $M6$ are the probabilities of land one, land two, ..., land six matching, and $NM1$, $NM2$, ..., $NM6$ are the probabilities that land one, land two, ..., land six do not match. However, to compute this probability, we would need to know the alignment of the two sets of lands. Fortunately, the consistent rotation of the bullet allows this to be done. For instance, if we knew that land 1 of bullet 1 matches to land 4 of bullet 2, then we immediately know that land 2 of bullet 1 matches to land 5 of bullet 2, land 3 of bullet 1 matches to land 6 of bullet 2, etc. Hence, we can take look across six diagonals of the $6 \otimes 6$ matrix containing match probabilities. Table 3.8 gives an example of the matrix of matching probabilities between two sets of six lands from bullets that are known matches. The matching diagonal is clear based on the high probabilities (cell $(1, 3)$, cell $(2, 4)$, cell $(3, 5)$, etc.) although it can be seen that one of the six comparisons has a relatively lower matching probability. This procedure is based on the Sequence Average Maximum (SAM) by Sensorfar (2017) in their bullet matching software application `SensoMatch`.

Table 3.8: Matrix of matching probabilities between two sets of six lands from bullets that are known matches.

profile1_id	45604	46104	46601	47069	47600	48069
42594	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
43063	0.0000	0.0000	0.0000	1.0000	0.0067	0.0000
43581	0.0000	0.0000	0.0000	0.0000	0.8433	0.0000
44211	0.0000	0.0000	0.0133	0.0000	0.0000	0.6700
44568	1.0000	0.0000	0.0033	0.0000	0.0000	0.0000
45070	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000

We derive a score by computing the bullet matching probability on each set of six matrix diagonals using the previously defined formula, under the assumption of land to land indepndence. Finally,

we take the maximum score obtained out of the six results as the final matching score for a bullet pair. After doing so, we can plot the scores for known matches and known non-matches separately. Figure 3.41 provides the distribution of matching scores for known matches compared to known non-matches. It can be seen that the known matches all have scores of around 100%, while no non-match achieves a score of above 30%, and hence this procedure provides perfect discrimination between all pairs of bullets between and within the two Hamby datasets.

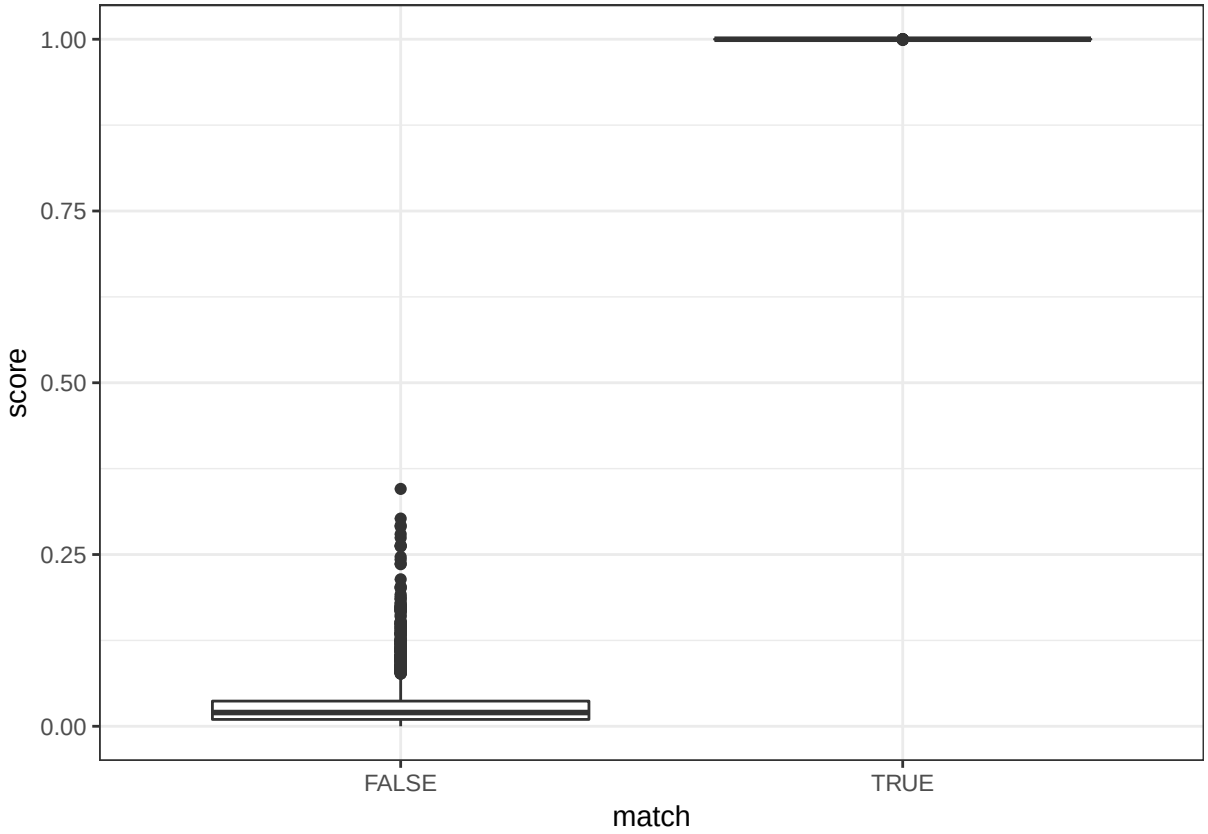


Figure 3.41: Distribution of matching scores for known matches compared to known non-matches. It can be seen that the known matches all have scores of around 100%, while no non-match achieves a score of above 30%.

On the other hand, flipping this procedure around by assuming that a match occurs if and only if all six lands match does not discriminate quite as well, as can be seen in Figure 3.42. Every known bullet non-match achieves a score of about zero, but so do about 15 known bullet matches. This method performs more poorly because our algorithm has a bigger problem with false negatives compared to false positives. Multiplying the probabilities together compounds

the issue of false negatives and leads to some misidentification of matching bullets, compared with the previous procedure.

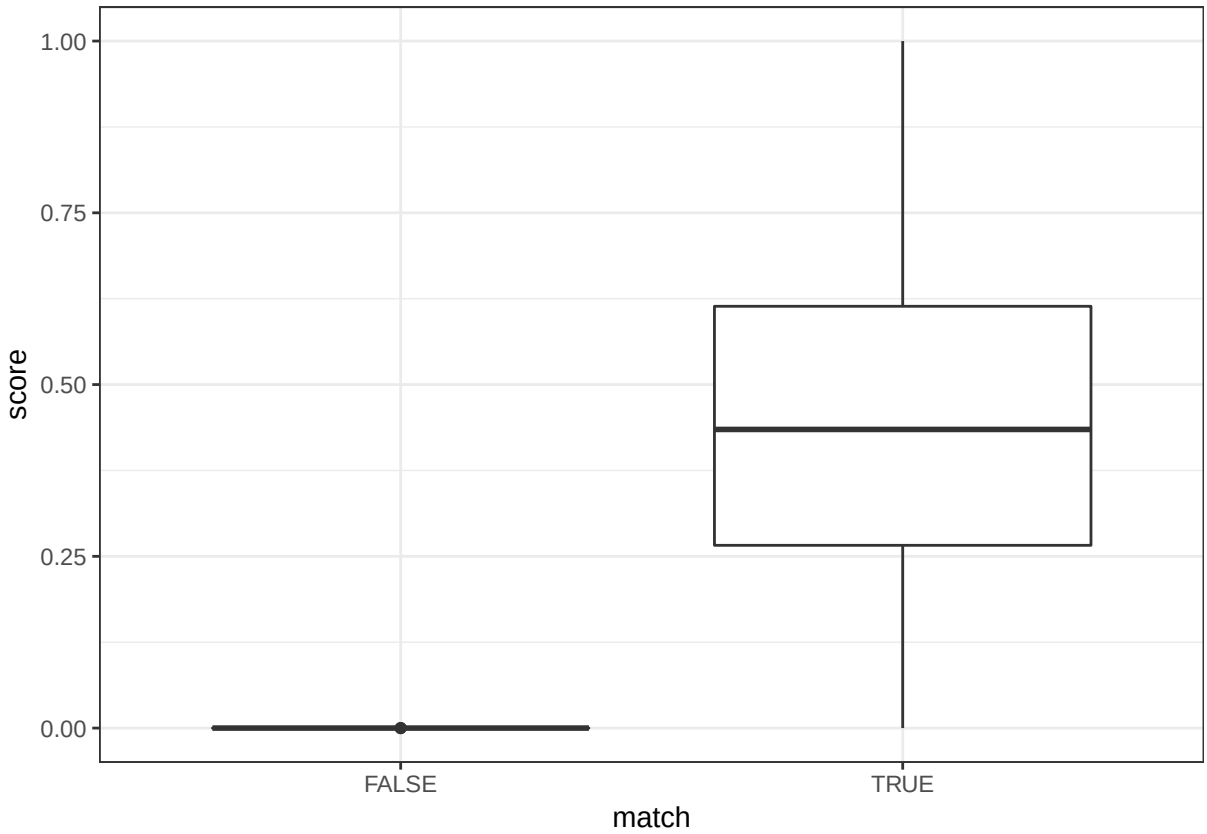


Figure 3.42: Distribution of matching scores for known matches compared to known non-matches when assuming a match occurs if and only if all six lands match. Now, it can be seen that the known non-matches all have scores of around 0, while most though not all known matches achieve much higher scores.

Taking a hybrid of these two approaches, we can average the probabilities along the diagonal rather than multiplying. Doing so yields the distributions shown in Figure 3.43. Now, we once again differentiate the two groups extremely well with no known non-match achieving a score above 10%, while no known match achieves a score below 40%.

One more approach to the bullet matching problem would be to exploit the SAM procedure on individual features. For each diagonal in the $6 \otimes 6$ matrix, we can compute an average value for each feature in our model. This will yield six sets of feature values for all six diagonals. We can then feed all six sets of features into the random forest in order to obtain a matching probability for each, taking the highest resulting probability to hopefully locate the diagonal which provides

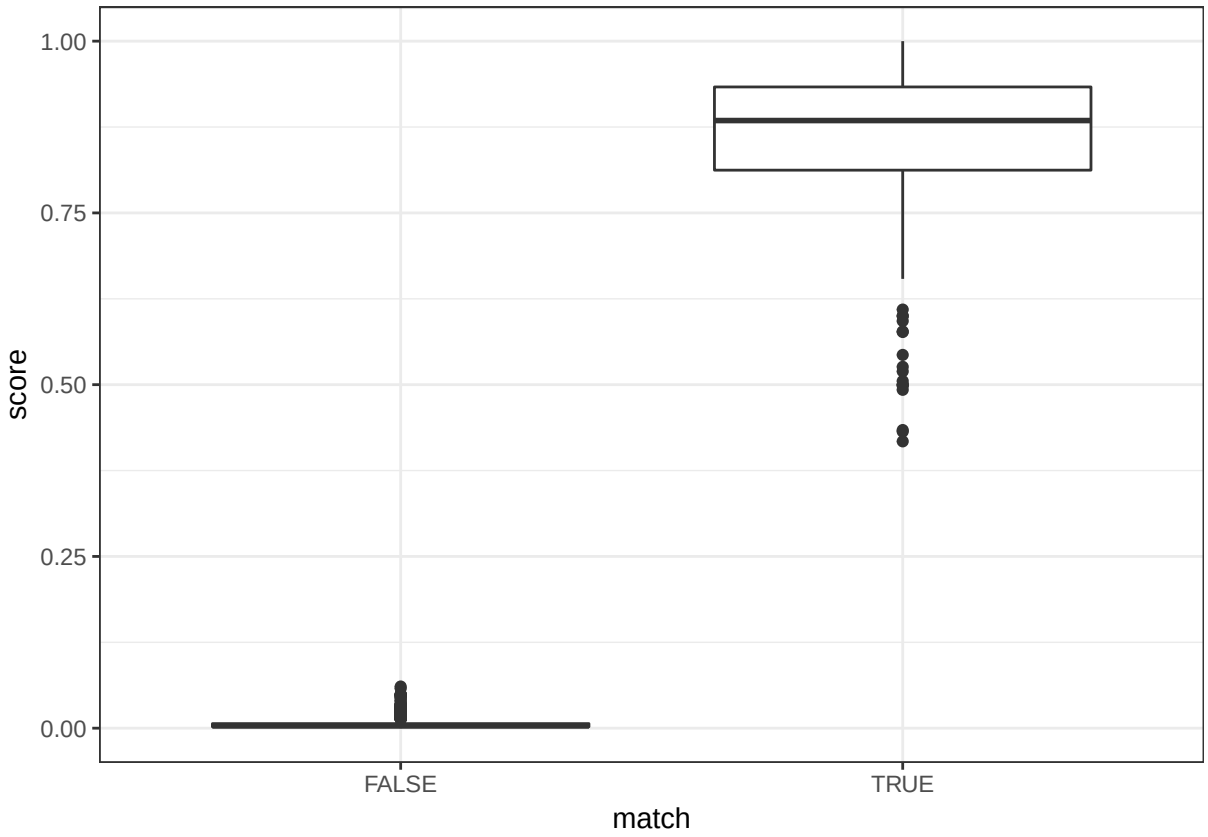


Figure 3.43: Distribution of matching scores for known matches compared to known non-matches obtained by averaging the probabilities along the maximal diagonal. Once again, it can be seen that the known non-matches all have scores of around 0, but the known matches this time are well separated, with nearly all achieving scores above .5.

the land to land alignment. Figure 3.44 provides boxplots of the matching scores using this procedure. It can be seen that while this procedure does discriminate well, it yields some false negatives (matching bullets that our forest identifies as a non-match).

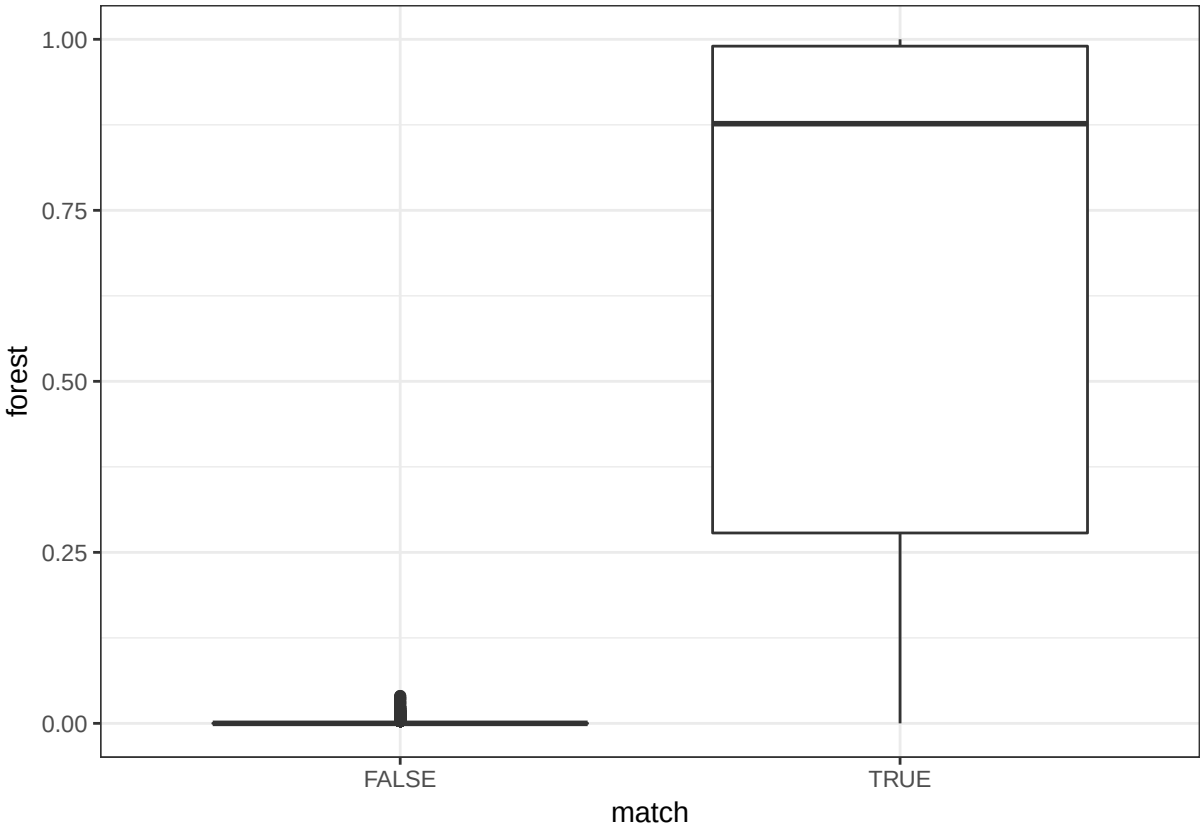


Figure 3.44: Distribution of matching scores using a SAM procedure on the feature values for known matches compared to known non-matches.

3.6 Conclusion

In this paper, we have introduced a set of robust features which can be used to train bullet matching models. We've used these features to train a random forest and assess its accuracy. In doing so, we noted strong evidence of operator effects being present in terms of the differences in the microscope scans.

While these effects were clear, the way in which this should be accounted for is less clear, however. In the ideal case, bullets fired from a particular gun barrel should yield surface scans

that are of identical quality and properties, regardless of the operator performing the scan. To achieve this, rigorous standards may need to be put in place with regards to the alignment of the bullet under the objective, and the procedure used to scan the bullet surface. Such procedures will require further investigation in order to lay out. For instance, due to the stark difference between the ideal cross section across two studies, procedures may need to dictate the margin from the edge of the objective at which the bullet can be placed.

Some first steps towards addressing the issue of degraded bullet lands was also taken. As suspected, the algorithm performance declines as a function of the amount of degradation. However, there is a relatively clear separation around about 50%. If 50% of the land or more is recovered, the algorithm performance is much stronger than values below that.

As before, further generalization and analysis of these algorithms are a bit limited by a lack of data. The assessment has still been performed on a controlled test set. The degraded land simulation itself may not represent entirely realistic scenarios of recovering fragmented bullets. However, as more data is collected, the model can be continually updated and retrained in order to improve its performance and handle more obscure cases other than the idealized versions we've currently been working with.

3.7 Appendix

Table 3.9: Results for the Kolmogrov-Smirnov distributional test.

set1	set2	feature	matchtest	matchd	nonmatchtest	nonmatchd
H252_H252	H252_H44	ccf	< 0.0001	0.2723	0.0001	0.0189
H252_H252	H252_H44	cms	< 0.0001	0.1751	< 0.0001	0.0245
H252_H252	H252_H44	D	< 0.0001	0.2567	< 0.0001	0.1049
H252_H252	H252_H44	matches	< 0.0001	0.1933	< 0.0001	0.0327
H252_H252	H252_H44	mismatches	< 0.0001	0.2015	0.3537	0.0079
H252_H252	H252_H44	overlap	0.0492	0.0984	< 0.0001	0.0276
H252_H252	H252_H44	rough_cor	< 0.0001	0.2647	< 0.0001	0.0970
H252_H252	H252_H44	sum_peaks	0.0008	0.1426	0.0015	0.0162
H252_H252	H44_H44	ccf	< 0.0001	0.2160	< 0.0001	0.0257
H252_H252	H44_H44	cms	< 0.0001	0.2515	< 0.0001	0.0467
H252_H252	H44_H44	D	< 0.0001	0.2342	< 0.0001	0.1946
H252_H252	H44_H44	matches	< 0.0001	0.2770	< 0.0001	0.0713
H252_H252	H44_H44	mismatches	< 0.0001	0.2505	0.0414	0.0138
H252_H252	H44_H44	overlap	0.2432	0.0906	< 0.0001	0.0408
H252_H252	H44_H44	rough_cor	< 0.0001	0.2242	< 0.0001	0.1718
H252_H252	H44_H44	sum_peaks	0.0001	0.1926	< 0.0001	0.0289
H252_H44	H44_H44	ccf	0.1149	0.0883	< 0.0001	0.0259
H252_H44	H44_H44	cms	0.111	0.0888	< 0.0001	0.0262
H252_H44	H44_H44	D	0.2923	0.0724	< 0.0001	0.0906
H252_H44	H44_H44	matches	0.0603	0.0977	< 0.0001	0.0423
H252_H44	H44_H44	mismatches	0.3301	0.0700	0.1633	0.0096
H252_H44	H44_H44	overlap	0.8231	0.0465	0.0001	0.0190
H252_H44	H44_H44	rough_cor	0.2671	0.0741	< 0.0001	0.0769
H252_H44	H44_H44	sum_peaks	0.047	0.1011	0.006	0.0147

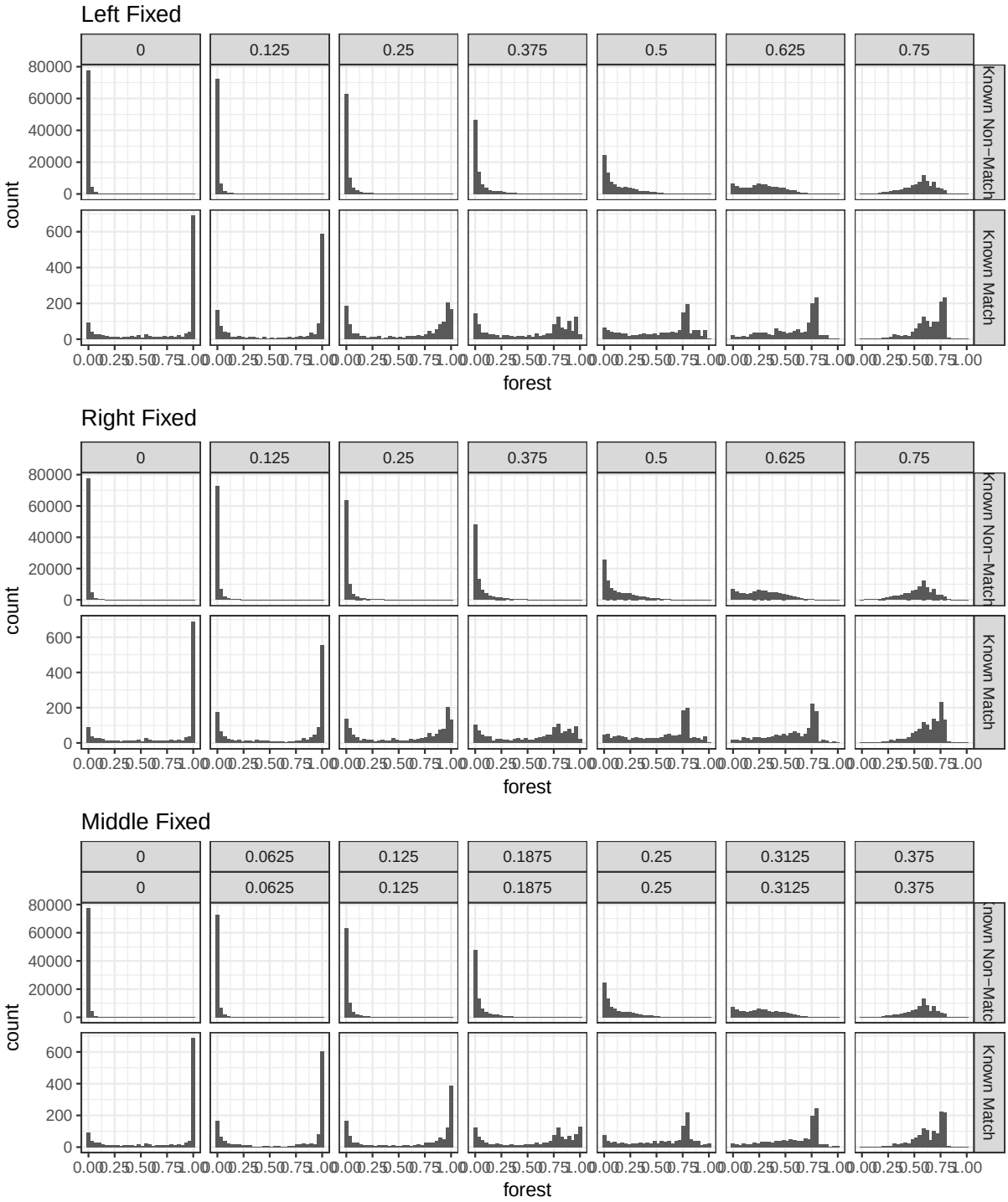


Figure 3.45: Histograms of matching probability, facetted by the degradation level and known match versus known non-match.

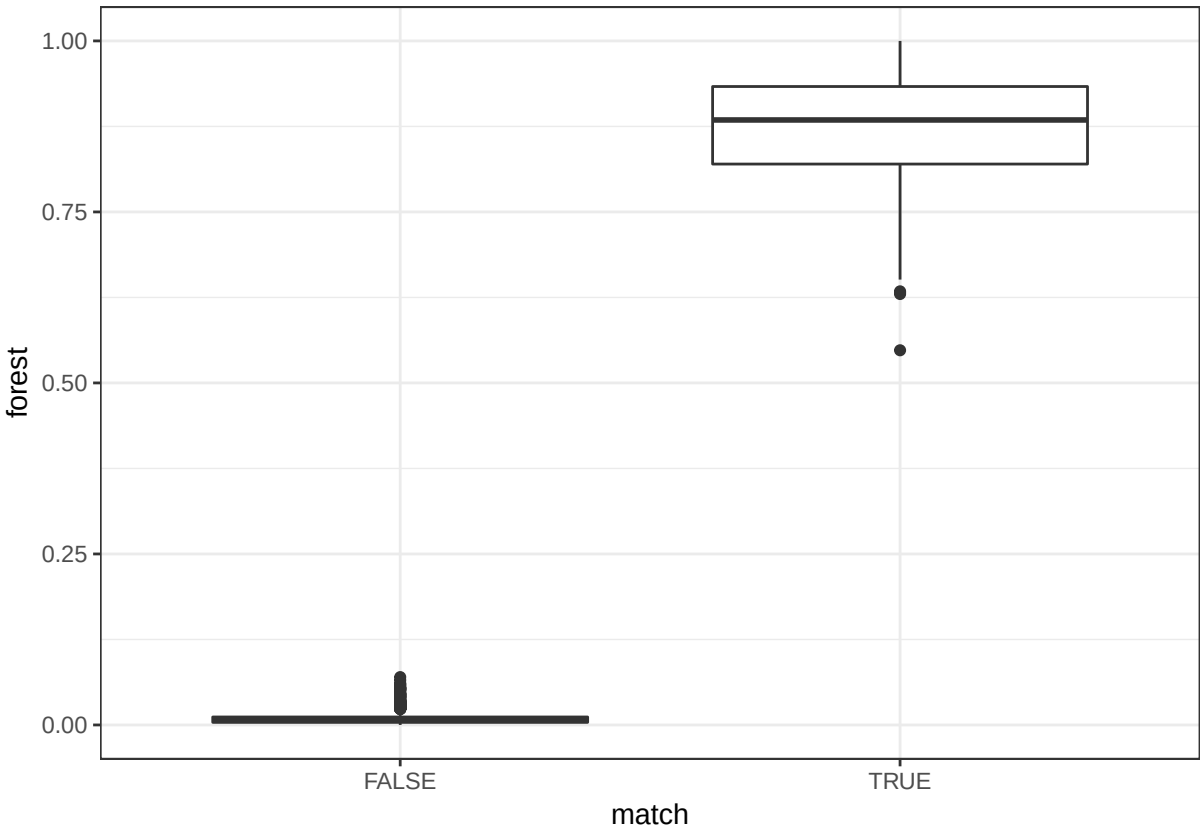


Figure 3.46: Score distributions for the naive approach to bullet matching, for known matches and known non-matches.

CHAPTER 4. A MODERN BULLET MATCHING DATABASE AND WEB APPLICATION

A paper to be submitted to the **Journal of Forensic Science**.

Eric Hare, Heike Hofmann, Alicia Carriquiry

Abstract

Bullet matching is a process used to determine whether two bullets have been fired from the same gun barrel. While traditionally a manual process performed by trained forensic examiners, recent work has been done to add statistical validity and objectivity to the procedure. In this paper, we build upon the algorithms explored in Automatic Matching of Bullet Lands by describing a database structure which tracks the parameters used through the course of the algorithm's run, and allows for the seamless replication of results by researchers, adding a layer of reproducibility to the bullet matching process. Finally, we describe two web applications, one intended for use by forensic examiners, and one intended for use by developers of bullet matching algorithms, which utilize the database to allow bullet matching to be done more efficiently and seamlessly.

4.1 Background

The need for advancements in terms of scientific objectivity and reproducibility of forensic methods is well known. Note, for example, the recent report by the President's Council of Advisors on Science on Technology (PCAST) (Advisors on Science and Technology 2016). The report references a number of areas of common practice in the field in which subjectivity is far too common, including but not limited to fingerprint analysis, bitemark analysis and firearms analysis.

Work has been done to address these concerns and add objectivity to these procedures. In the case of firearms analysis, the focus of this paper, some examples include Vorburger et al. (2011), Chu et al. (2013), and Riva and Champod (2014). Our own work, "Automatic Matching of Bullet Lands", describes procedures used to produce an estimate of the probability of a match between two bullet lands. It does so by deriving a number of features, some from the literature and some original, and computing these features on pairs of reference bullets from the NIST Ballistics Toolmark Research Database (NBTRD). The algorithms used are published as open-source R code available in the package `bulletr` (Hofmann and Hare 2016). In spite of these steps towards transparency, however, the process of duplicated and assessing the performance of the algorithm in hopes of improving predictive accuracy was cumbersome in a number of ways:

1. Doing so requires specialized statistical software (specifically, R and associated R packages)
2. Computing statistics on all pairs of bullet lands is a time consuming process even on high-powered machines (on the order of several days)
3. Updates to our `bulletr` package, or any package dependencies of `bulletr`, may change the results such that our findings are not completely reproducible even if each step is correctly followed

In this paper, we add a new layer of reproducibility to the algorithms to allow for forensic scientists, statisticians, and other interested researchers to duplicate and iterate on the results

in a seamless fashion. We do so by introducing a new database structure that supplements the NIST database by storing all necessary parameters and intermediate results needed to arrive at a matching probability between two bullet lands. We describe the structure from a technical perspective, and then describe the front-end and back-end application structure which utilizes the database to provide results to the researchers. Finally, we provide a case study analysis on features of bullet land pairs as an example of the capabiilities of this structure when it is leveraged. Reproducible R (R Core Team 2016) code used to access the database is provided as a convenience to future researchers working in this domain.

4.2 Database Structure

Figure 4.47 displays the database schema along with links between the relevant id columns. This structure provides the necessary links between the raw input data, and the processed signatures used to compute features and ultimately provide matching probabilities between pairs of lands. This diagram will be explained in depth in this section.

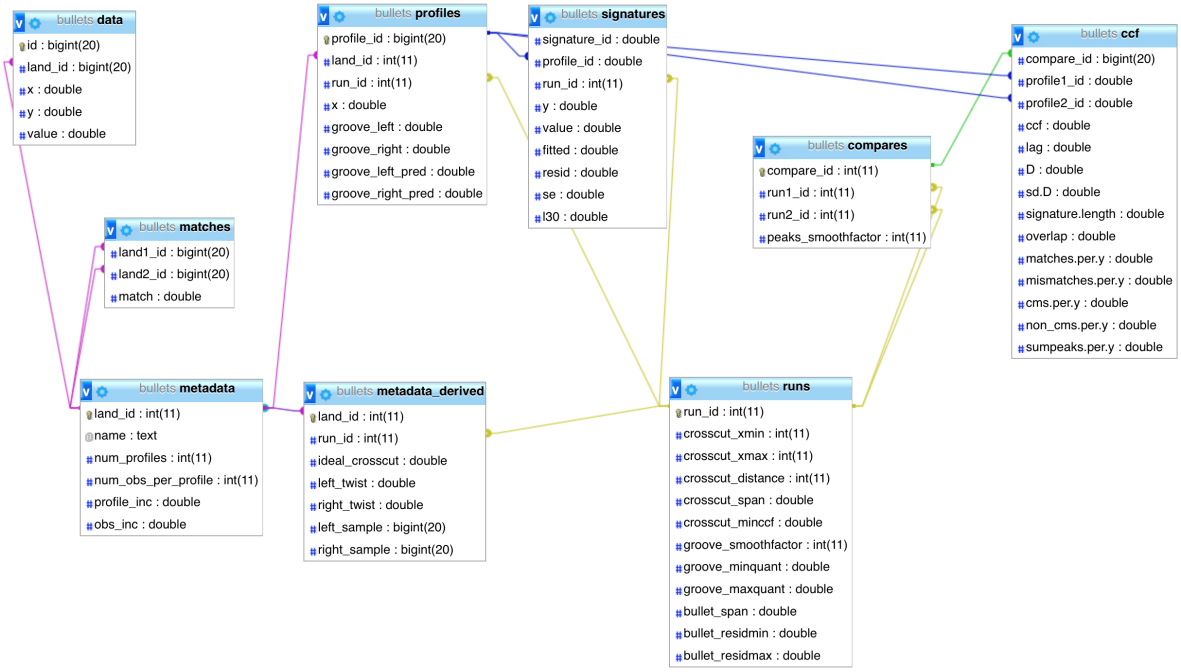


Figure 4.47: A schematic of the database.

We can connect to the database using the `dplyr` package.

```
library(xtable)
library(RMySQL)
library(tidyverse)
library(bulletr)
library(gridExtra)
library(randomForest)
library(caret)
library(broom)
library(matrixcalc)


dbname <- "bullets"
user <- "buser"
password <- readLines("buser_pass.txt")
host <- "127.0.0.1"
port <- 3306


my_db <- src_mysql(dbname, host, port, user, password)
```

The remainder of this section will walk through the most relevant database tables, and include reproducible R code for accessing, parsing, and displaying the data.

4.2.1 Data

The data table is essentially a mirror of the bullets stored in the NBTRD. It currently includes a long-form version of the two Hamby bullet sets (Set 252 and Set 44) (Hamby, Brundage, and Thorpe 2009) and the Cary Persistence study (Wong, n.d.). A sample of 20 rows of this table can be seen in Table 4.10. The land id column identifies the bullet land under consideration.

The x coordinate is the location along the shorter axis, while the y coordinate is along the longer axis. The value column represents the height of the bullet at that particular location.

```
my_data <- tbl(my_db, "data")


# Get Hamby Barrel 1 Bullet 1 Land 3
result <- my_data %>% filter(land_id == 39, !is.na(value)) %>%
    arrange(x, y) %>% head(n = 20) %>% as.data.frame
```

Table 4.10: A sample of 20 rows of the data table. The land id column identifies the bullet land under consideration. The x coordinate is the location along the shorter axis, while the y coordinate is along the longer axis. The value column represents the height of the bullet at that particular location.

	id	land_id	x	y	value
1	32992516	39	0	34.38	-2.07
2	32993018	39	0	35.94	-1.57
3	32993520	39	0	37.50	-1.26
4	32994022	39	0	39.06	-1.36
5	32996532	39	0	46.88	1.91
6	32997034	39	0	48.44	2.34
7	32997536	39	0	50.00	2.56
8	32998038	39	0	51.56	3.04
9	32998540	39	0	53.12	3.26
10	32999042	39	0	54.69	4.24
11	32999544	39	0	56.25	4.11
12	33000046	39	0	57.81	4.29
13	33000548	39	0	59.38	4.78
14	33001050	39	0	60.94	5.24
15	33001552	39	0	62.50	5.98
16	33002054	39	0	64.06	5.89
17	33004062	39	0	70.31	8.22
18	33004564	39	0	71.88	8.55
19	33005066	39	0	73.44	8.66
20	33005568	39	0	75.00	8.84

In comparison to a regular x3p file, this data table is less space efficient. An x3p file uses a surface matrix of dimension (x, y) where the value of each (x_i, y_j) is the height of the bullet at $x = i$ and $y = j$. Our expanded version of the format turns each cell into a single row x_i, y_j, z_{ij}. Where z_{ij} where z_{ij} is the height of the bullet at $x = i$, $y = j$. Thus, for a bullet with 500 x

values and 1572 y values (as in land id 39), an x3p file uses a matrix of size 786000, while we use a data frame of 786000 rows and 3 columns, storing $3x$ as much information. While certainly less space efficient, the format allows for easy querying of specific lands or specific profiles from the database.

4.2.2 Metadata

As shown in Table 4.11, the metadata table includes one row for each unique bullet land. The name of the bullet is derived from the file path, as provided by the NBTRD. Several other parameters are given, including the number of profiles, (x values), number of observations per profile (y values), and the increments of each in micrometers. (in this case, one x unit is equivalent to 1.5625 micrometers). Note that this table includes parameters that are derived solely from the properties of the data itself. Hence, the previous data table in conjunction with metadata forms the information needed to generate x3p files. Conversely, x3p files can be used to regenerate the data contained in both these tables.

```
my_metadata <- tbl(my_db, "metadata")


result <- my_metadata %>% filter(land_id >= 61, land_id <=
    72) %>% as.data.frame
```

4.2.3 Metadata Derived

Similarly to the metadata table, the metadata_derived table (Sampled in Table 4.12) includes one row per bullet land. The difference is that the columns of this table were derived by our algorithm rather than properties of the data. The run_id, which will be discussed in more depth later, indicates the algorithm run that yielded the following derived parameters.

```
my_metadata_derived <- tbl(my_db, "metadata_derived")
```

Table 4.11: A sample of 12 rows of the metadata table. The land id column identifies the bullet land under consideration. The name of the bullet is derived from the file path, as provided by the NBTRD. Several other parameters are given, including the number of profiles, (x values), number of observations per profile (y values), and the increments of each in micrometers. (in this case, one x unit is equivalent to 1.5625 micrometers)

land_id	study	barrel	bullet	land	num_profiles	num_obs_per_profile
61	Hamby252	2	1	1	501	1573
62	Hamby252	2	1	2	502	1582
63	Hamby252	2	1	3	502	1461
64	Hamby252	2	1	4	502	1568
65	Hamby252	2	1	5	500	1572
66	Hamby252	2	1	6	502	1569
67	Hamby252	2	2	1	502	1508
68	Hamby252	2	2	2	502	1533
69	Hamby252	2	2	3	502	1568
70	Hamby252	2	2	4	502	1565
71	Hamby252	2	2	5	502	1591
72	Hamby252	2	2	6	500	1596

```
result <- my_metadata_derived %>% filter(land_id >= 61,
    land_id <= 72, run_id == 3) %>% as.data.frame
```

A number of derived parameters are given for a particular land and a particular run:

1. **ideal_crosscut** - The location of the ideal cross section (or ideal x coordinate) at which to extract a profile, as given by Hare, Hofmann, Carriquiry (2017).
2. **left_twist** - The calculated twist of the scan as determined by the left shoulder.
3. **right_twist** - The calculated twist of the scan as determined by the right shoulder.
4. **left_sample** - The number of samples used to compute the left twist.
5. **right_sample** - The number of samples used to compute the right twist.

4.2.4 Profiles

Table 4.13 displays 10 profiles from land id 39, using the first 20 x coordinate values. Profiles are defined by properties of the grooves or shoulders. Hence, given this information, we can post-process the data table to extract particular profiles. For instance, we can use profile_id

Table 4.12: A sample of 12 rows of the metadata_derived table. Once again, a land id identifies a particular bullet land. The run id, which will be discussed in more depth later, indicates the algorithm run that yielded the following derived parameters.

land_id	run_id	ideal_crosscut	left_twist	right_twist	left_sample	right_sample
61	3	150	0.1270	0.1461	406	406
62	3	150	0.1390	0.1952	404	404
63	3	150	0.1421	0.1013	406	406
64	3	300	0.1064	0.1180	408	408
65	3	100	0.1143	0.1166	419	419
66	3	125	0.1215	0.1408	409	409
67	3	150	0.1288	0.0363	401	401
68	3	150	0.1049	0.1444	409	409
69	3	125	0.1346	0.1255	400	400
70	3	175	0.1301	0.1277	408	408
71	3	150	0.1139	0.1278	406	406
72	3	175	0.1197	0.1214	409	409

32448, which is the profile obtained by extracting at x = 100 for land_id 39. Using properties of the grooves as determined by our algorithm, we can extract the the profile with the shoulders and grooves removed.

```
my_profiles <- tbl(my_db, "profiles")


result <- my_profiles %>% filter(land_id == 39, x > 92,
    x < 107) %>% select(-groove_left_pred, -groove_right_pred) %>%
    as.data.frame
```

Figure 4.48 displays the profile obtained by extracting land id 39 at x = 100. Dashed vertical lines indicate the location of the shoulders. Within the bounds of the dashed line, the profiles that are relevant for bullet matching are obtained.

```
myprof <- filter(result, x == 100)


land39 <- my_data %>% filter(land_id == 39, x == 100) %>%
    as.data.frame
```

Table 4.13: A sample of 10 rows of the profiles table. A profile id is uniquely identified by a land id, a run id, and an x value.

profile_id	land_id	run_id	x	groove_left	groove_right
19325	39	1	92.19	179.69	2128.12
19326	39	1	93.75	179.69	2129.69
19327	39	1	95.31	179.69	2129.69
19328	39	1	96.88	181.25	2129.69
19329	39	1	98.44	181.25	2129.69
19330	39	1	100.00	181.25	2129.69
19331	39	1	101.56	181.25	2129.69
19332	39	1	103.12	182.81	2131.25
19333	39	1	104.69	182.81	2131.25
19334	39	1	106.25	182.81	2129.69

```
ggplot(data = land39, aes(x = y, y = value)) + geom_line() +
    geom_vline(xintercept = myprof$groove_left, linetype = 2) +
    geom_vline(xintercept = myprof$groove_right, linetype = 2) +
    theme_bw()
```

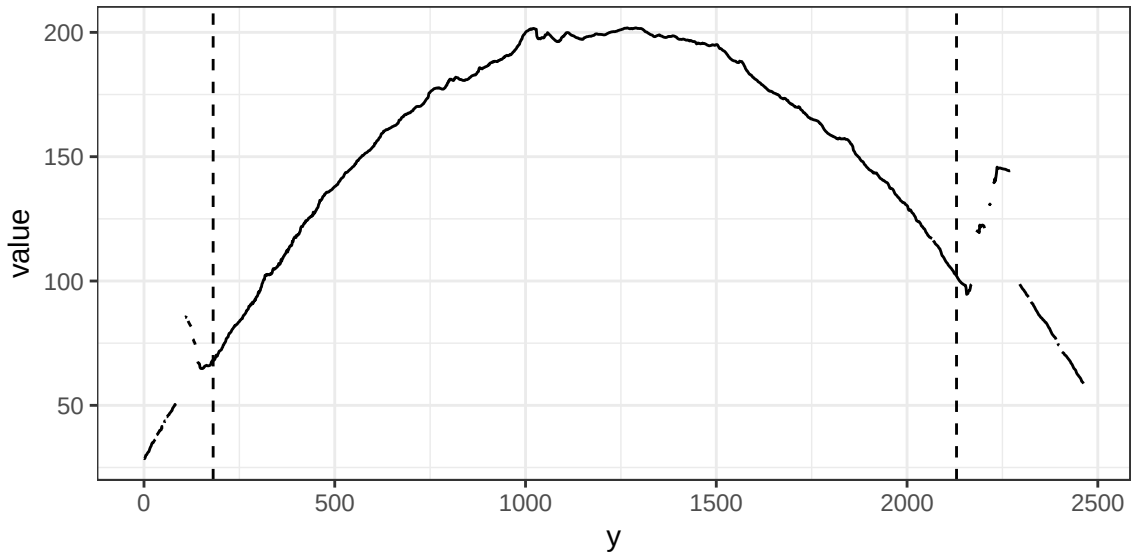


Figure 4.48: The profile obtained by extracting land id 39 at x = 100. Dashed vertical lines indicate the location of the shoulders. Within the bounds of the dashed line, the profiles that are relevant for bullet matching are obtained.

4.2.5 Signatures

The land signature represents the processed data that is ultimately used for matching. In our case, a land signature represents the smoothed and processed residuals obtained from fitting a Locally Weighted Scatterplot Smoothing Regression (LOESS) to the profiles from above (Cleveland 1979). Figure 4.49 displays the signature obtained by processing the profile of land id 39 at x = 100. It can be seen that the signature represents an attempt at reducing a bullet land to the peaks and valleys that represent striations, by removing the global structure of the bullet that dominates the view of the profile.

```
my_signatures <- tbl(my_db, "signatures")


result <- my_signatures %>% filter(profile_id == myprof$profile_id,
    run_id == 1) %>% as.data.frame


ggplot(data = result, aes(x = y, y = l30)) + geom_line() +
    theme_bw()
```

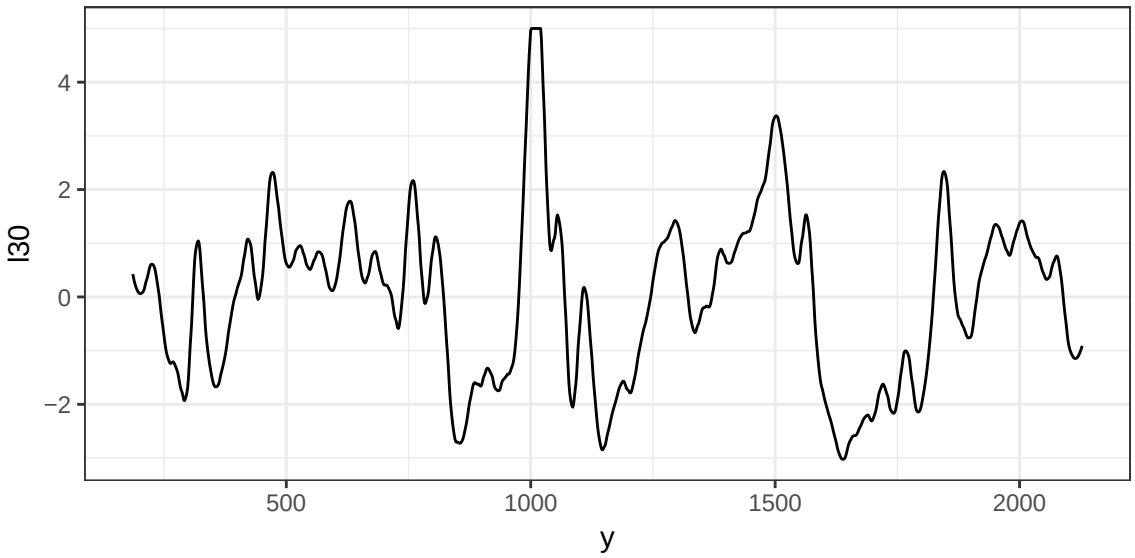


Figure 4.49: The signature obtained by processing the profile of land id 39 at x = 100.

4.2.6 CCF

The CCF table contains features computed on cross-comparisons between different signatures. The name is a bit of a misnomer; the ccf, or cross-correlation function, is only one of the features. Table 4.14 displays a subset of the derived features for a comparison of the derived profile for land id 39, from above, with six other land profiles. This land's known match is the fourth row in the table, and the features immediately stand out as more pronounced, including a ccf above 90% and a number of matches far exceeding the other comparisons.

```
my_ccf <- tbl(my_db, "ccf")


result <- my_ccf %>% filter(profile1_id == myprof$profile_id,
    compare_id == 4) %>% select(profile1_id, profile2_id,
    ccf, rough_cor, D, overlap, matches, cms, sum_peaks) %>%
    collect() %>% slice(5:10) %>% as.data.frame
```

Table 4.14: A subset of the derived features for a comparison of the derived profile for land id 39, from above, with six other land profiles. This land's known match is the fourth row in the table, and the features immediately stand out as more pronounced, including a ccf above .9 and a number of matches far exceeding the other comparisons.

profile1_id	profile2_id	ccf	rough_cor	D	overlap	matches	cms	sum_peaks
19330	21894	0.383	0.170	0.003	0.956	4.324	1.622	9.237
19330	22364	0.368	-0.226	0.004	0.914	2.261	1.696	3.295
19330	22882	0.195	-0.130	0.003	0.877	1.768	1.179	3.034
19330	23366	0.902	0.822	0.001	0.996	9.412	5.752	12.776
19330	23852	0.280	-0.019	0.002	0.899	1.727	0.576	2.675
19330	24306	0.182	-0.130	0.003	0.943	2.747	1.648	4.988

4.3 Web Applications

We now turn our attention to two applications which build upon the previously described database. These applications were designed to be web-based, easy to use, interactive applications which supplement forensic examiners and forensic scientists in either performing a bullet matching routine, or participating in bullet matching research.

4.3.1 Front-End

Figure 4.50 displays the first page of the application. Using the database, the application populates two lists along the left-hand side, allowing selection of any of the bullet lands currently stored in the database. By default, two lands that are known to match, from bullets fired from the first Hamby 252 barrel, are used. The forensic examiner can also upload their own bullet lands, in the x3p format (OpenFMC 2014), and the application will use those lands rather than lands stored in the database.

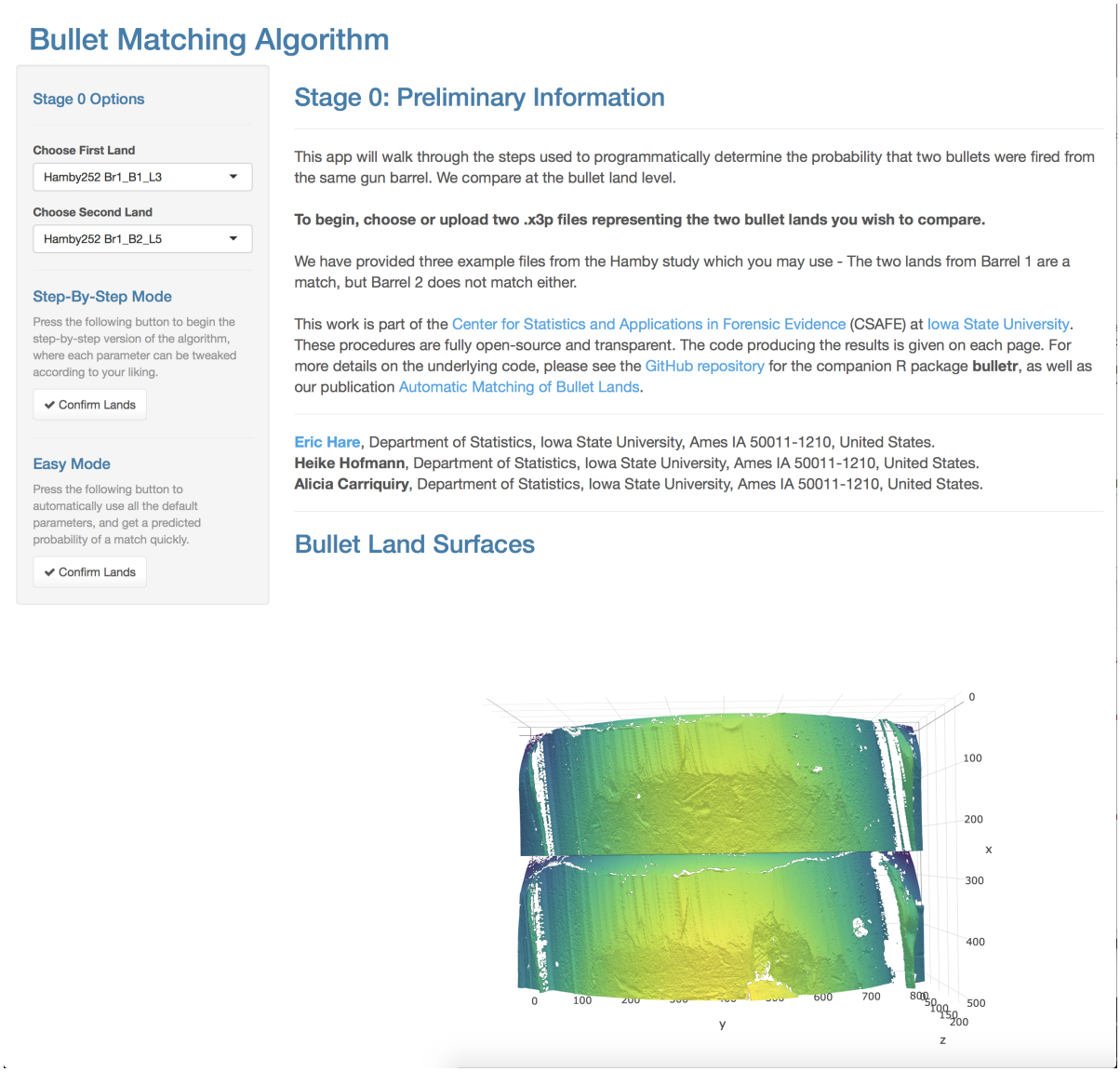


Figure 4.50: User Interface for the front-end bullet matching algorithm.

The first page is called Stage 0. Stage 0 involves selection of the bullet lands, and provides some preliminary information. Beneath the land selection are two buttoms to either select "Step-By-Step Mode" or "Easy Mode". This document will focus on "Step-By-Step Mode", which enforces that the user interact with the process of the algorithm. By contrast, "Easy Mode" will choose parameters of the algorithm, such as the location to take a cross section, or the level of smoothing, automatically based on the procedures described in Automatic Matching of Bullet Lands.

Using the `plotly` package (Sievert et al., n.d.), the two lands are rendered in a 3D viewing framework. This framework allows panning, rotation, zooming, and other features to aid in the manual and visual inspection of the bullet land surfaces. These surface renderings will be displayed on each page of the application so that they can be used to help inform some parameter choices.

When the forensic examiner has chosen the bullet lands and read the information on Stage 0, they can choose the "Confirm Lands" button underneath "Step-By-Step Mode" to begin the matching process. The application then moves onto Stage 1, "Finding a Stable Region". In this stage, the goal is to select the coordinate of the ideal cross-section of each land. Using the algorithm described in Hare, Hofmann, Carriquiry (2016), the application attempts to select what it believes is the ideal cross-section, and provides those for each land as the default choice. When satisfied with the choice, the forensic examiner then can select "Confirm Coordinates" to continue to Stage 2.

Stage 2 involves automatic detection and removal of the grooves. A portion of the application at this stage is shown in Figure 4.51. In the top left, sliders are available representing the regions at which to extract the profile. While typically the default choice works well, some profiles may have unusual patterns that make automatic groove detection inaccurate. In these cases, the sliders can be adjusted as necessary to fine tune the locations. The vertical blue lines plotted show the groove location coordinates overlaid with the profile. Performing this step allows the

LOESS fit to be impacted only by the curved structure of the bullet land itself, and not by the dominating grooves.

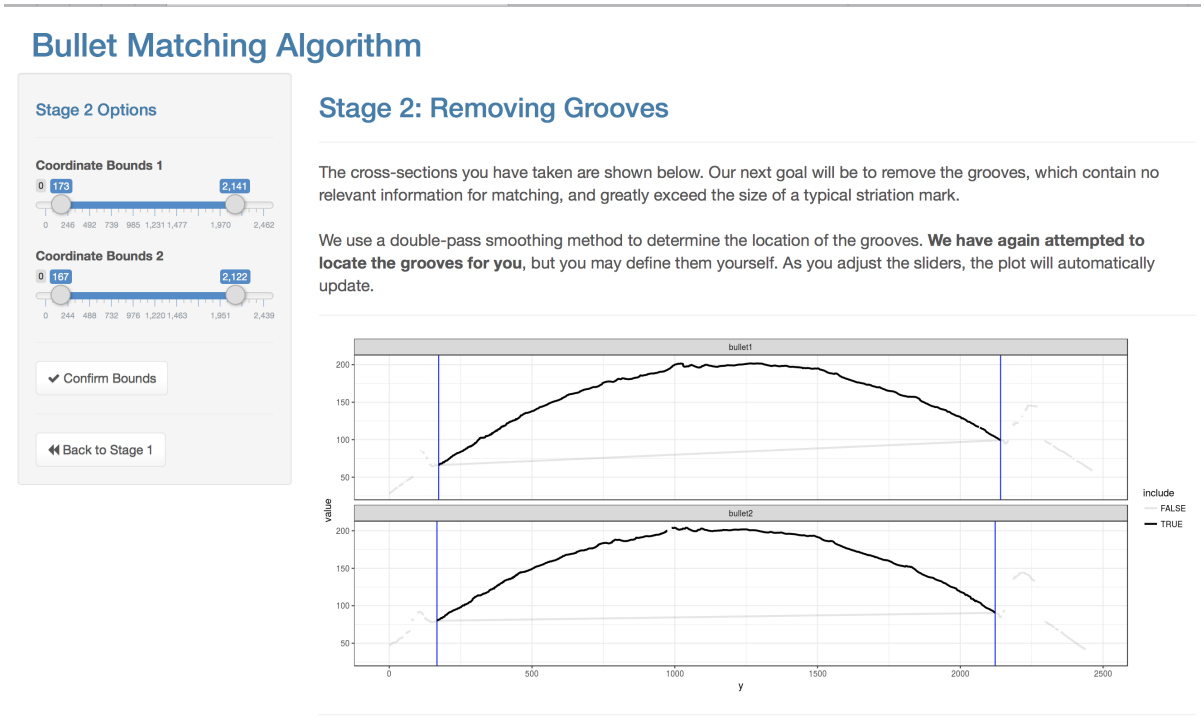


Figure 4.51: Stage 2 of the front-end bullet matching application. In this stage, the grooves are automatically detected and removed from the profile of the land.

After groove detection, the application moves into Stage 3 (Figure 4.52). The application automatically fits a LOESS regression with a span of 3%, and allows the forensic examiner to adjust the span as desired to control the level of smoothing. The fits for each land, along with the extracted residuals (the land signature), are displayed.

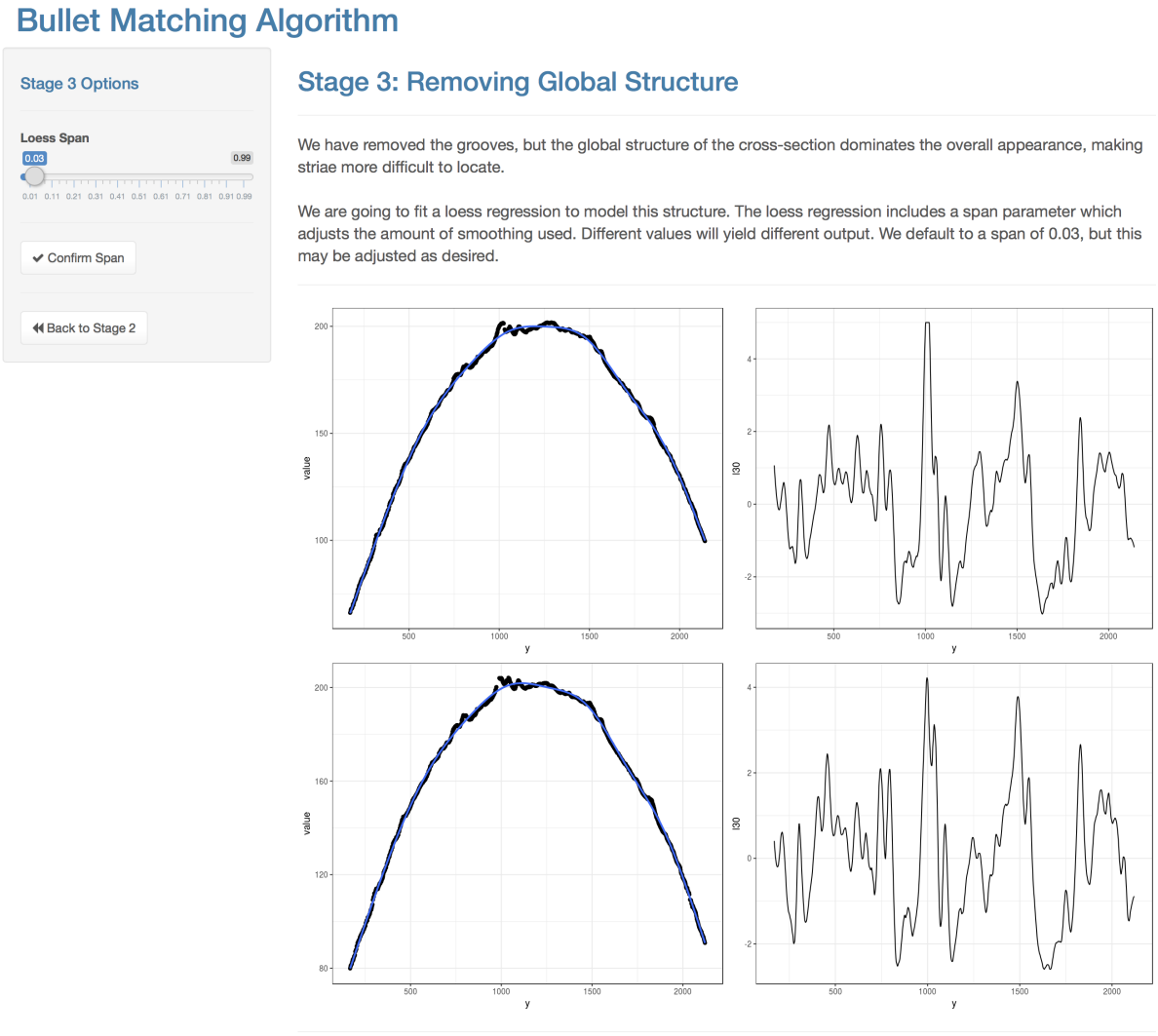


Figure 4.52: Stage 3 of the front-end bullet matching application. In this stage, a LOESS regression is fit to the resulting profile (with grooves removed), and the smoothed residuals or the "signature" of the bullet lands are extracted.

Stage 4 (Figure 4.53) is an alignment stage. Using the ccf, the two signatures are aligned automatically for purposes of extracting features. As in prior stages, the amount of lag can be adjusted manually if the automatic choice is not ideal.

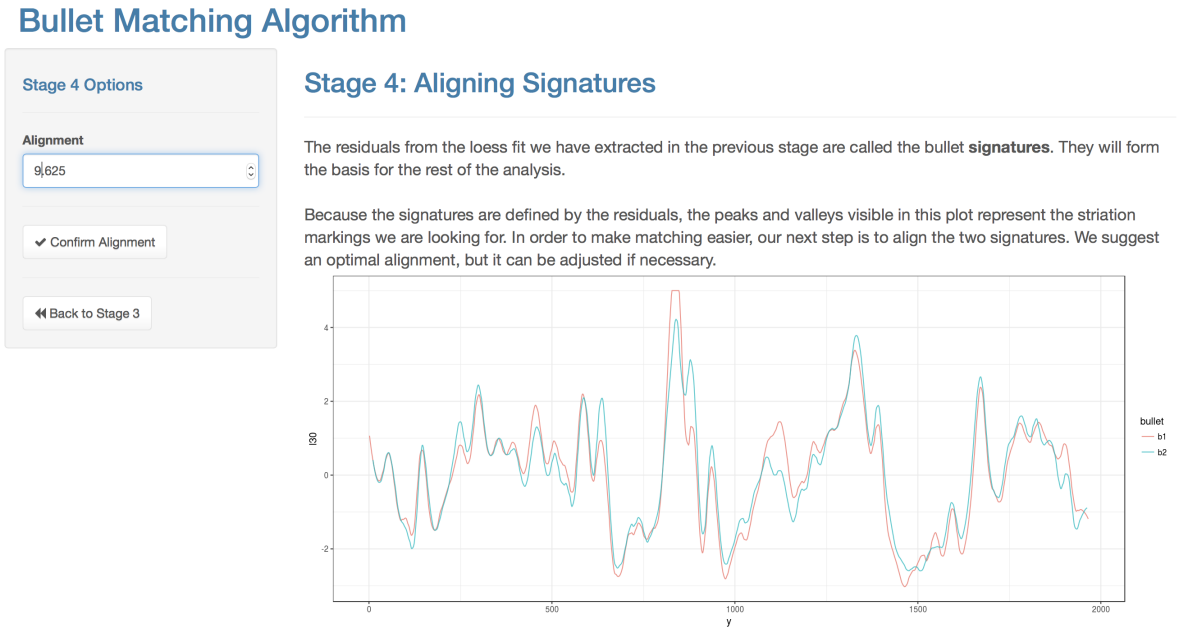


Figure 4.53: Stage 4 of the front-end bullet matching application. Here, the two land signatures are automatically aligned.

The final stage, Stage 5, involves using the aligned signatures in order to detect peaks and valleys. By smoothing over the signatures of each land, locations in which the derivative is equal to zero can be detected. Figure 4.54 displays the application at this stage. Note that the level of smoothing, called the "Smoothing Factor", can be adjusted as desired. The detected peaks and valleys in the aligned signatures are indicated by red and blue vertical lines, respectively.

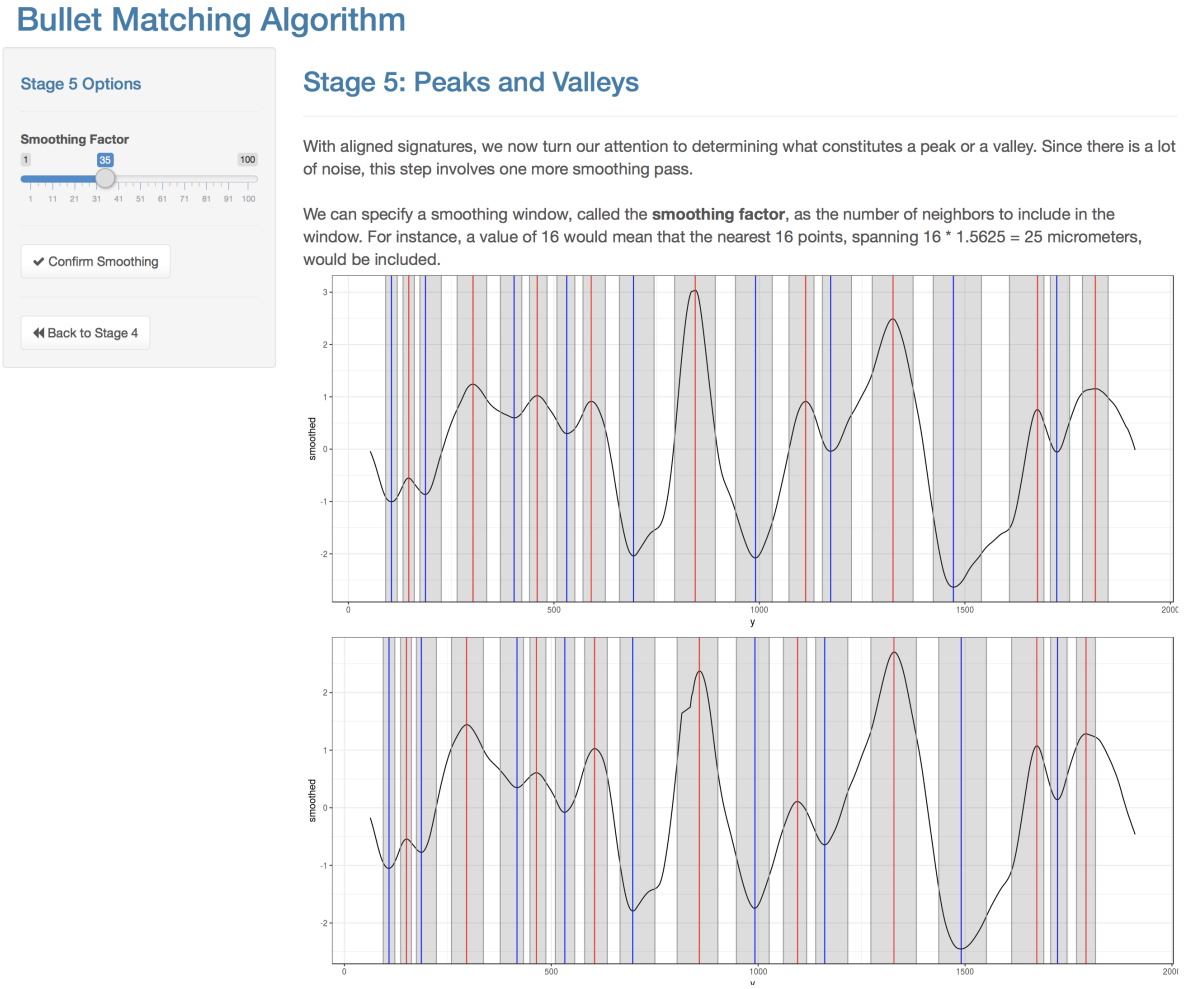
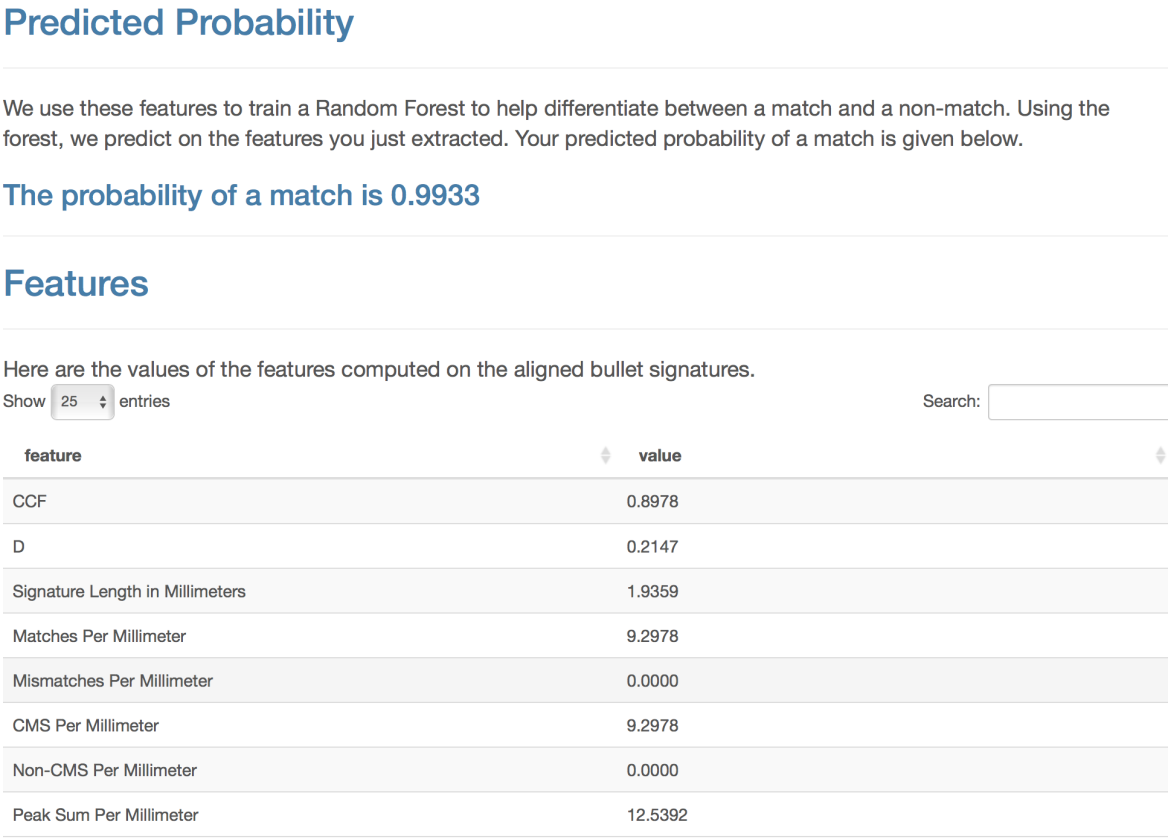


Figure 4.54: Stage 5 of the front-end bullet matching application, where peaks and valleys are detected in the smoothed and aligned signatures.

Finally, all stages are completed, and the resulting report is generated. A subset of the report is shown in Figure 4.55. In particular, note that the probability of a match based on the trained random forest is given, along with the derived features. Although not shown in the figure, the report also includes the results and parameters of all previous stages, for reproducibility. In this way, by printing the results, a step-by-step trace of each stage of the algorithm can be performed.



Predicted Probability

We use these features to train a Random Forest to help differentiate between a match and a non-match. Using the forest, we predict on the features you just extracted. Your predicted probability of a match is given below.

The probability of a match is 0.9933

Features

Here are the values of the features computed on the aligned bullet signatures.

Show 25 entries Search:

feature	value
CCF	0.8978
D	0.2147
Signature Length in Millimeters	1.9359
Matches Per Millimeter	9.2978
Mismatches Per Millimeter	0.0000
CMS Per Millimeter	9.2978
Non-CMS Per Millimeter	0.0000
Peak Sum Per Millimeter	12.5392

Figure 4.55: A portion of the final output of the bullet matching application.

4.3.2 Back-End

The back-end application, shown in Figure 4.56, stands in contrast to the front-end application by being primarily intended for researches looking to improve the matching performance of the algorithm. Like the front-end app, it uses the database to generate a list of all the bullet lands available. Unlike the front-end app, however, it allows the selection of only one land. Once it is selected, the ideal cross section will be displayed on the left.

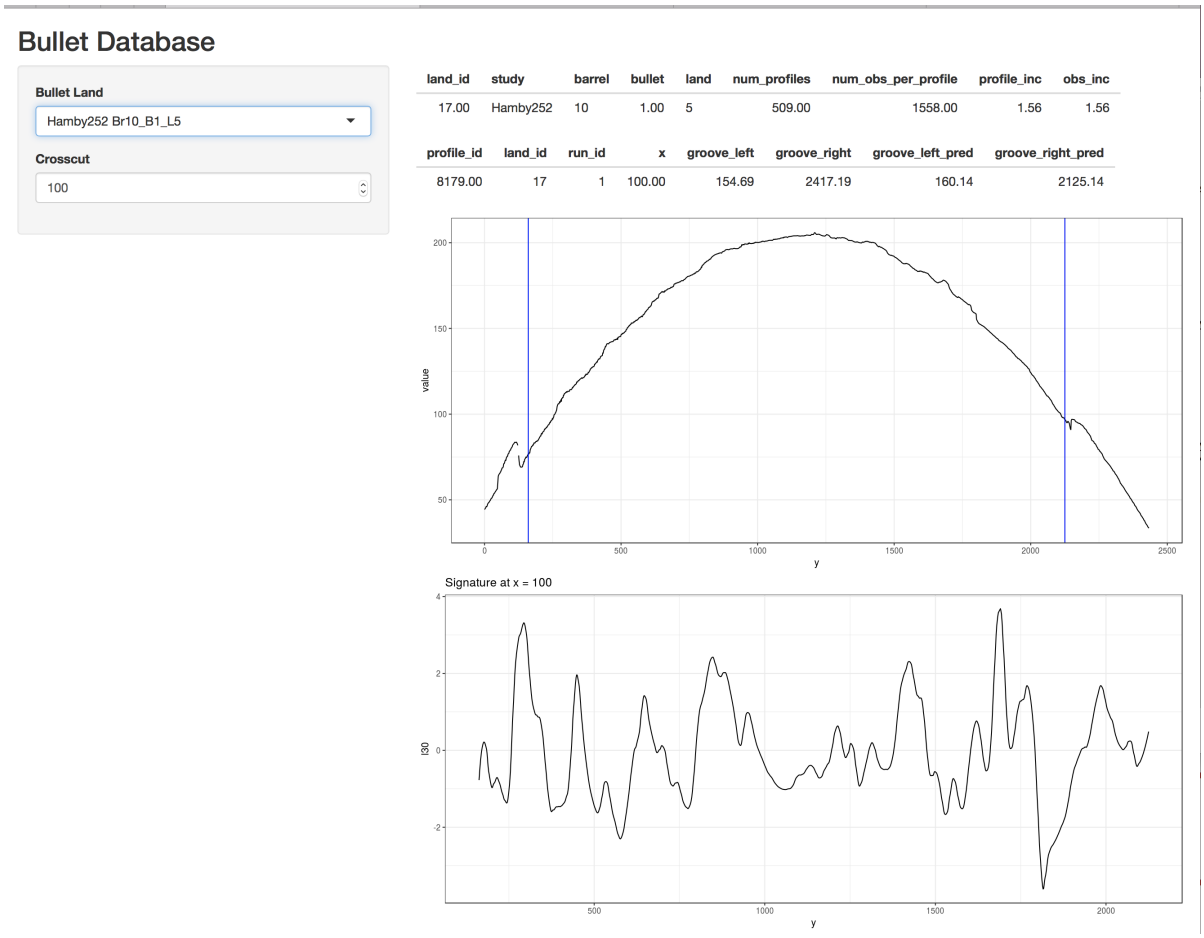


Figure 4.56: User Interface for the back-end bullet matching algorithm.

On the right hand side, there will be two tables. The first table is the metadata for that particular land, which comes directly from the original x3p file. The second table is information on the profile based on the selected cross-section. Beneath this table, the profile with the detected grooves overlaid is shown. Finally, beneath that, the signature derived from this profile is shown.

While functionally simple, this application allows for the assessment of the generated signatures. In the course of tuning parameters and optimizing performance, we can examine signatures that may have issues, for instance if the grooves were not properly detected or a poor cross section was taken.

4.4 Conclusion

In this paper, we have introduced and described a formal database housing raw bullet data, and the results of each processing stage of our bullet matching algorithm. Because this database is openly accessible, and all parameters are tracked, this allows for researchers to more easily use the results in order to iterate on components of the algorithm in hopes of improving the matching performance. Furthermore, use of the database will ensure that algorithms built upon it will automatically update as new data is provided, so that over time results can improve.

The two web applications both serve important purposes in the process of bullet matching research. The front-end application serves as an entry point to the algorithms and the features which a non-programmer can utilize. By simply uploading the surface scans of two bullet lands, a predicted matching probability can be obtained in a few seconds. Perhaps more importantly, values of the features can also be obtained, and the bullet lands themselves can be rotated and viewed in a 3D plotting framework. This application can act as a supplement to classical comparison methods, which are traditionally done manually under a comparison microscope.

On the other hand, the back-end application is intended for researchers who intend to develop code to improve upon the algorithms. It allows an assessment of where the algorithm is having the most issue distinguishing matches from non-matches, and can be used to quickly implement improvements as needed.

CHAPTER 5. FUTURE WORK

There are a number of areas demanding further research and exploration to turn this procedure into something that can and should be used in a criminal justice setting. Ultimately, the biggest limitation of the algorithms described is in terms of the reference database. At the time of this writing, the Hamby and Cary studies remain the only publicly available scanned bullets in the 3D surface format. From Hamby, the two sets comprise 70 bullets that were fired from 10 barrels. The Cary study is a single barrel. In other words, there are 11 total gun barrels used to generate the results derived in this document. Naturally, this is a limiting factor in terms of the out-of-sample performance. While the initial results are promising, confidence can not and should not be too high in the procedure before the reference database contains a wide variety of gun barrels, fired under a number of real world scenarios, and the algorithms have been tested using bullets recovered from crime scenes. In particular, the simulation of degraded data may or may not match real world degradation scenarios, and it would be far more useful to assess on a controlled set of true degraded data. The good news is that as bullets are added to the database, the algorithms can be rerun seamlessly to generate updated results.

Another area to explore is the extensive amount of parameters available for tweaking. While we have attempted to use cross-validation to optimize a number of these, including the optimal smoothing factor, there are also other parameters that have been fixed. Ideally, we could tune the parameters for each cross-comparison of lands. This would naturally require significant computational power and time, but would likely lead to more accurate predictions compared with the fixed parameter choices. Even if fixed values are chosen, they can also be optimized such that the fixed values lead to the best accuracy given cross-validation.

While we have documented the presence of operator or scan effects, we didn't make many recommendations in terms of how these should be accounted for. It is simple to verbalize a solution - rigorous procedures for scanning, consistent across labs and operators. However, in practice this is challenging. We could attempt to quantify the scan parameters across operators and adjust for them in some manner, but if the scans themselves are too low quality, this is unlikely to be effective. On the other hand, the model themselves could include an operator or microscope component that is included to handle this automatically. At this point, with two sets of scans, we have too little data to work with, but this will certainly need to be explored as time goes on.

One area that may limit the adoption of automated or semi-automated computer-based algorithms in the forensic community is the desire to still perform manual alignment and matching using expert training. To supplement the examiners, a full-featured 3D viewer that operates as a comparison microscope is needed. We have a very basic version of this, allowing the zooming, panning, and rotating of the bullet lands. But fine-grained alignment of the striations is challenging or impossible in the current framework.

Finally, there should be more exploration into a generative model for land signatures. If an accurate generative model could be created, we could begin to create likelihood ratios that truly assess the odds that a particular bullet was fired from a gun, in a way that could have significant sway in court. Unfortunately, given the wide variety of firing conditions, barrels, ammunition, etc., such a generative model could be very challenging to create. We have chosen the route of avoiding this issue, and instead using an empirical model based on the features generated on land-to-land comparisons. Still, this is an area of research that will likely be a significant pursuit going forward.

BIBLIOGRAPHY

Advisors on Science, President's Council of, and Technology. 2016. "Report on Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods." https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

AFTE Criteria for Identification Committee. 1992. "Theory of Identification, Range Striae Comparison Reports and Modified Glossary Definitions – Afte Criteria for Identification Committee Report." *AFTE Journal* 24 (2): 336–40.

Bachrach, B. 2002. "Development of a 3D-based automated firearms evidence comparison system." *J. Forensic Sci.* 47 (6): 1253–64.

Biasotti, Alfred A. 1959. "A Statistical Study of the Individual Characteristics of Fired Bullets." *Journal of Forensic Sciences* 4 (1): 34–50.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.

Breiman, Leo, Jerry H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. Wadsworth, Inc.

Chang, Winston, Joe Cheng, JJ Allaire, Yihui Xie, and Jonathan McPherson. 2015. *Shiny: Web Application Framework for R*. http://CRAN.R-project.org/package=shiny.

Chu, W., J. Song, T. Vorburger, R. Thompson, and R. Silver. 2011. "Selecting Valid Correlation

Areas for Automated Bullet Identification System Based on Striation Detection." *Journal of Research of the National Institute of Standards and Technology* 116 (3): 649.

Chu, W., J. Song, T. Vorburger, J. Yen, S. Ballou, and B. Bachrach. 2010. "Pilot study of automated bullet signature identification based on topography measurements and correlations." *J. Forensic Sci.* 55 (2): 341–47.

Chu, Wei, Robert M Thompson, John Song, and Theodore V Vorburger. 2013. "Automatic identification of bullet signatures based on consecutive matching striae (CMS) criteria." *Forensic Science International* 231 (1–3): 137–41.

Clarkson, James A, and C Raymond Adams. 1933. "On Definitions of Bounded Variation for Functions of Two Variables." *Transactions of the American Mathematical Society* 35 (4). JSTOR: 824–54.

Cleveland, William S. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74 (368). Taylor & Francis, Ltd.: 829–36. http://www.jstor.org/stable/2286407.

Committee to Assess the Feasibility, Accuracy and Technical Capability of a National Ballistics Database. 2008. *Ballistic Imaging*. Edited by Daniel L. Coork, Eugene S. Meieran, and Carol V. Petrie. Washington, DC: The National Academies Press. doi:{10.17226/12162}.

De Ceuster, J., and S. Dujardin. 2015. "The reference ballistic imaging database revisited." *Forensic Science International* 248 (March): 82–87.

De Kinder, J., F. Tulleners, and H. Thiebaut. 2004. "Reference ballistic imaging database performance." *Forensic Science International* 140 (2-3): 207–15.

Fellows, Ian. 2012. "Deducer: A Data Analysis Gui for R." *Journal of Statistical Software* 49 (8).

Fox, John. 2005. "The R Commander: A Basic-Statistics Graphical User Interface to R."

Journal of Statistical Software 14 (9).

Giannelli, Paul C. 2011. "Ballistics Evidence Under Fire." *Criminal Justice* 25 (4): 50–51.

Hamby, James E., David J. Brundage, and James W. Thorpe. 2009. "The Identification of Bullets Fired from 10 Consecutively Rifled 9mm Ruger Pistol Barrels: A Research Project Involving 507 Participants from 20 Countries." *AFTE Journal* 41 (2): 99–110.

Hare, E., H. Hofmann, and A. Carriquiry. 2016. "Automatic Matching of Bullet Lands." *ArXiv E-Prints*, January.

Hofmann, Heike, and Eric Hare. 2016. *Bulletr: Algorithms for Matching Bullet Lands*.

Jed Wing, Max Kuhn. Contributions from, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, et al. 2016. *Caret: Classification and Regression Training*. https://CRAN.R-project.org/package=caret.

Liaw, Andy, and Matthew Wiener. 2002. "Classification and Regression by RandomForest." *R News* 2 (3): 18–22. http://CRAN.R-project.org/doc/Rnews/.

Lock, Amy B, and Max D Morris. 2013. "Significance of Angle in the Statistical Comparison of Forensic Tool Marks." *Technometrics* 55 (4). Taylor & Francis: 548–61.

Ma, L., J. Song, E. Whitenton, A. Zheng, T. Vorburger, and J. Zhou. 2004. "NIST bullet signature measurement system for RM (Reference Material) 8240 standard bullets." *Journal of Forensic Sciences* 49 (4): 649–59.

Milborrow, Stephen. 2015. *Rpart.plot: Plot 'Rpart' Models: An Enhanced Version of 'Plot.rpart'*. http://CRAN.R-project.org/package=rpart.plot.

Miller, J. 1998. "Criteria for Identification of Toolmarks." *AFTE Journal* 30 (1): 15–61.

National Research Council. 2009. *Strengthening Forensic Science in the United States: A Path*

Forward. Washington, DC: The National Academies Press. doi:10.17226/12589.

Nichols, Ronald G. 1997. "Firearm and Toolmark Identification Criteria: A Review of the Literature." *Journal of Forensic Sciences* 42 (3): 466–74.

———. 2003a. "Consecutive Matching Striations (CMS): Its Definition, Study and Application in the Discipline of Firearms and Tool Mark Identification." *AFTE Journal* 35 (3): 298–306.

———. 2003b. "Firearm and Toolmark Identification Criteria: A Review of the Literature: Part Ii." *Journal of Forensic Sciences* 48 (2): 318–27.

OpenFMC. 2014. *X3pr: Read/Write Functionality for X3p Surface Metrology Format*.

Petraco, Nicholas, and Helen Chan. 2012. *Application of Machine Learning to Toolmarks: Statistically Based Methods for Impression Pattern Comparisons*. Mannheim, Germany: Bibliographisches Institut AG.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Riva, Fabiano, and Christophe Champod. 2014. "Automatic Comparison and Evaluation of Impressions Left by a Firearm on Fired Cartridge Cases." *Journal of Forensic Sciences* 59 (3): 637–47. doi:10.1111/1556-4029.12382.

Roberge, D., and A. Beauchamp. 2006. "The Use of Bullettrax-3D in a Study of Consecutively Manufactured Barrels." *AFTE Journal* 38 (2): 166–72.

RStudio Team. 2015. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc. http://www.rstudio.com/.

S-PLUS Programmer's Manual. 1992. Version 3.1. Seattle, WA, USA: StatSci, a Division of MathSoft, Inc.

Sensorfar. 2017. *SensoMATCH Bullet Comparison Software*.

Sievert, Carson, Chris Parmer, Toby Hocking, Scott Chamberlain, Karthik Ram, Marianne Corvellec, and Pedro Despouy. n.d. *Plotly: Create Interactive Web-Based Graphs via Plotly's Api*. http://CRAN.R-project.org/package=plotly.

Therneau, Terry, Beth Atkinson, and Brian Ripley. 2015. *Rpart: Recursive Partitioning and Regression Trees*. http://CRAN.R-project.org/package=rpart.

Tukey, John Wilder. 1977. *Exploratory Data Analysis.* Addison-Wesley.

Vorburger, T.V., J.-F. Song, W. Chu, L. Ma, S.H. Bui, A. Zheng, and T.B. Renegar. 2011. "Applications of Cross-Correlation Functions." *Wear* 271 (3–4): 529–33. doi:http://dx.doi.org/10.1016/j.wear.2010.03.030.

Wickham, Hadley. 2009. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. http://had.co.nz/ggplot2/book.

Wong, Cary. n.d. "Cary Persistence."

Xie, F., S. Xiao, L. Blunt, W. Zeng, and X. Jiang. 2009. "Automated Bullet-Identification System Based on Surface Topography Techniques." *Wear* 266 (5–6): 518–22. doi:http://dx.doi.org/10.1016/j.wear.2008.04.081.

Xie, Yihui. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. http://yihui.name/knitr/.